CEBS Microarray Analysis User's Guide

November, 2005 CEBS Microarray 1.6.1 Release

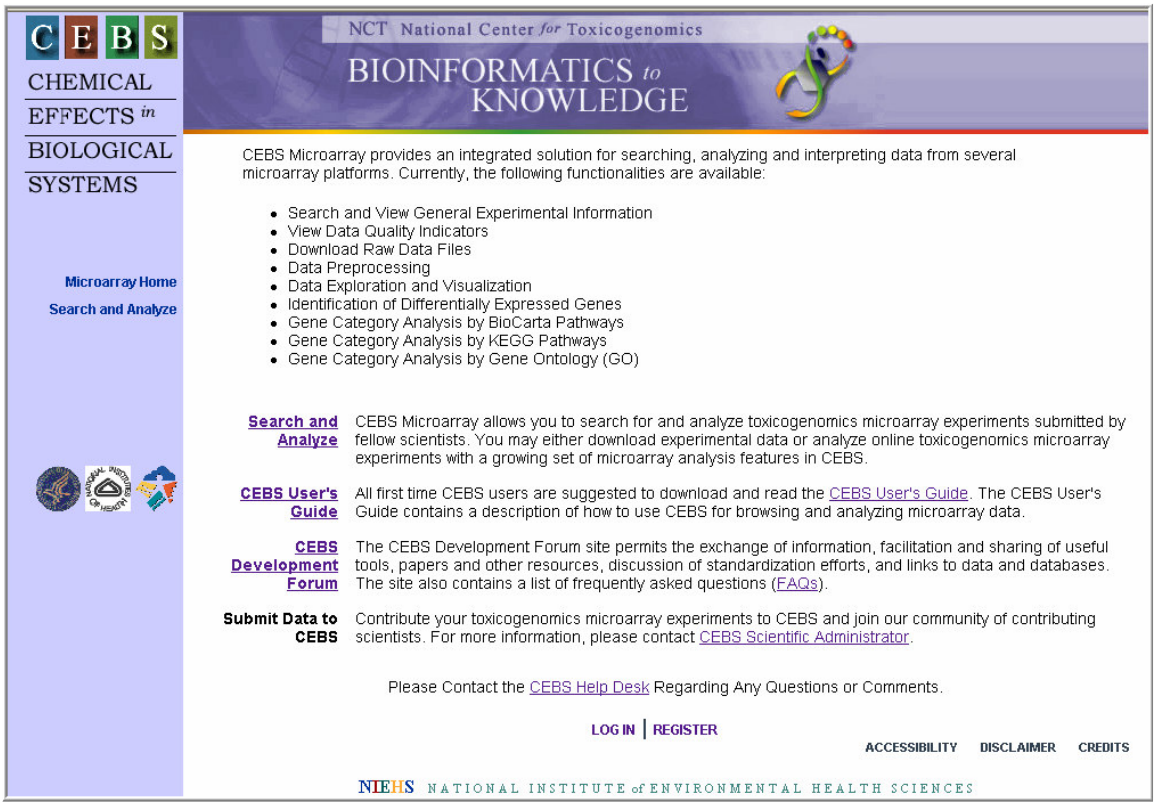
1. Introduction and General Overview

1.1. General Introduction to CEBS Microarray

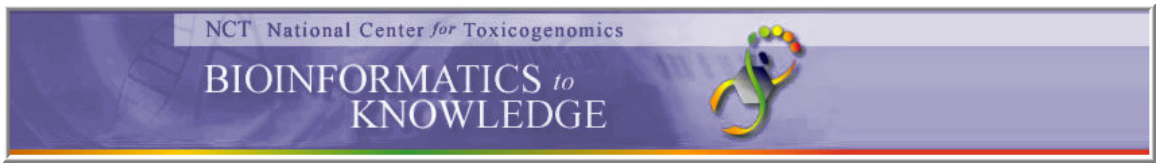
The Chemical Effects in Biological Systems (CEBS) is a knowledge base currently in development. An important component of CEBS is the CEBS Microarray Data Portal and Analysis Tool Suite that hosts microarray toxicogenomics data along with on-line analysis tools.

1.2. CEBS Common Page Components

Each web page in CEBS has a number of sections. These sections are found on almost every web page in CEBS.



1.2.1. Header



The Header is the banner at the top of each web page.

1.2.2. Left Menu

There are two text links and three icon links on the Left Menu.

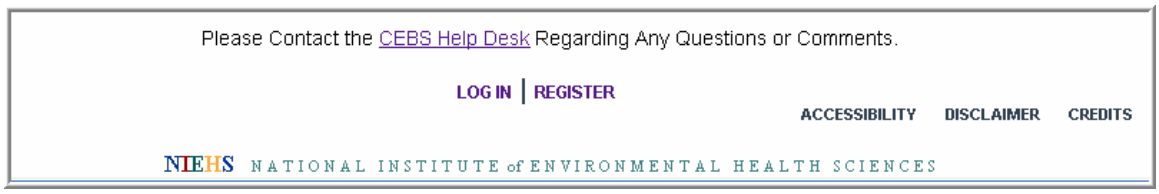
Text Links:

- Microarray Home: Link back to the microarray home page.
- Search and Analyze: Link to the Search Experiment Page

Icon Links:

- HHS: A link to www.hhs.gov.
- NIH: A link to www.nih.gov.
- NIEHS: A link to www.niehs.nih.gov

1.2.3. Footer



There are seven text links that can be found in the footer:

Text Links:

- LOG IN : login to CEBS
- REGISTER : register as a new user to CEBS
- LOGOUT: logout of CEBS
- CEBS Help Desk: email contact to request help
- Accessibility: Accessibility Information.
- Disclaimer: NIEHS Disclaimer of Endorsement
- Credits

Note: The Login and Register link will both be displayed together. If you have not logged into CEBS, these links will be displayed. If you have logged in, then the Logout link will be displayed.

1.2.4. Main Body

The main body of the page will vary greatly from page to page. Most screen shots found in this document will consist of only the main body of the page to allow the viewer to focus on the important aspects of the page.

1.2.5. Buttons

There are a number of buttons that can be found on most of the web pages.

Continue – Allows you to continue to the next page in the workflow scenario. In some cases, it will continue to the next page and submit the information from the current form to CEBS.

Submit – Allows you to submit the information from the current form to CEBS.

1.3. Definition of Experiment Visibility

The visibility of an experiment refers to which users are able to access experiment information and data. A user can specify the "visibility" of an experiment he/she is submitting to ensure that sensitive (i.e. pre-publication) data is only accessible by authorized users (for example, in user's consortium research group).

- Private – visible only to users with the same Principal Investigator, in other words, members of the same research group.
- Public – visible to anyone accessing the CEBS Microarray site.
- Group Restricted Access – visible to all users who are part of the Group

There are a number of different groups such as TRC Consortium. If you wish to add a group to CEBS, please send an email to cebsfeedback@list.niehs.nih.gov.

2. Accessing the Site

To access the site, open a browser and type http://cebs.niehs.nih.gov/microarray into the address bar. This will take you to the welcome page as shown in Figure 1 below.

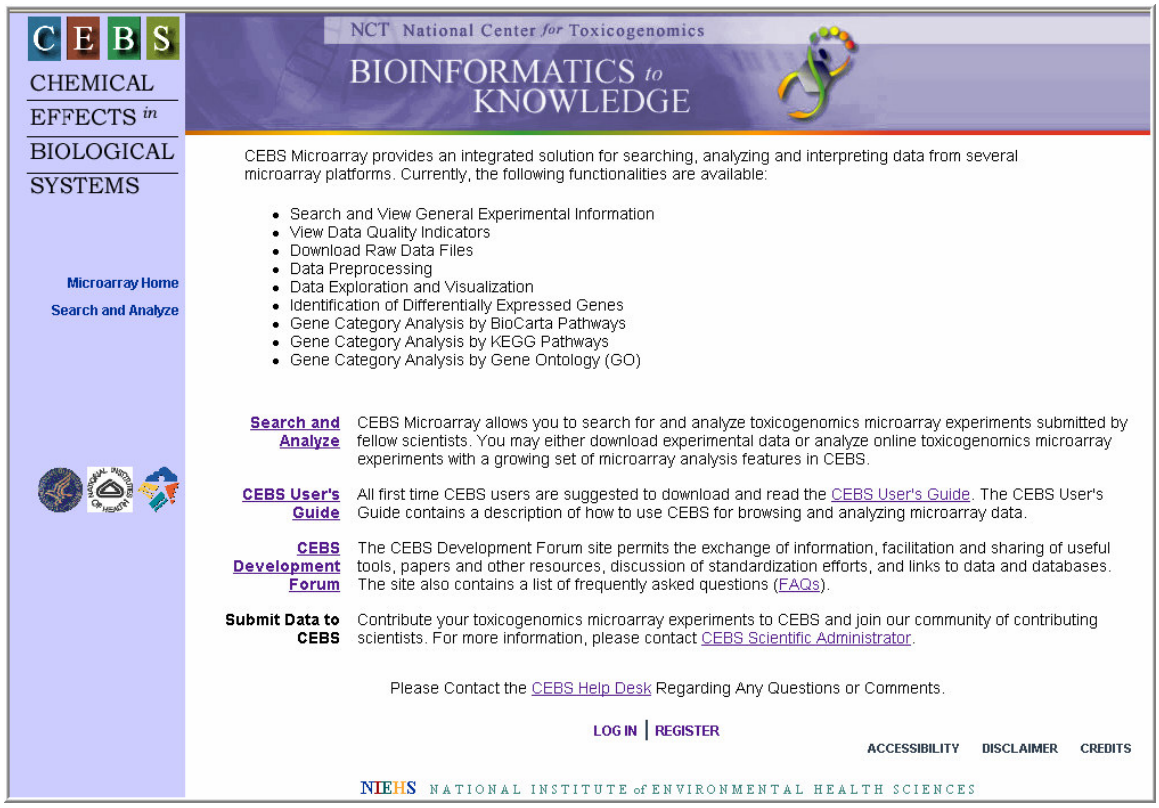


Figure 1 - CEBS Microarray Home Page

3. Registering a New CEBS Microarray Account

In order to submit an experiment, you must have a user account to sign into the application. Also, to search experiments with restricted access, or "visibility", it is necessary to have a CEBS Microarray user's account.

From the CEBS Microarray Home Page, click on the link – Register to access the CEBS Microarray account registration page

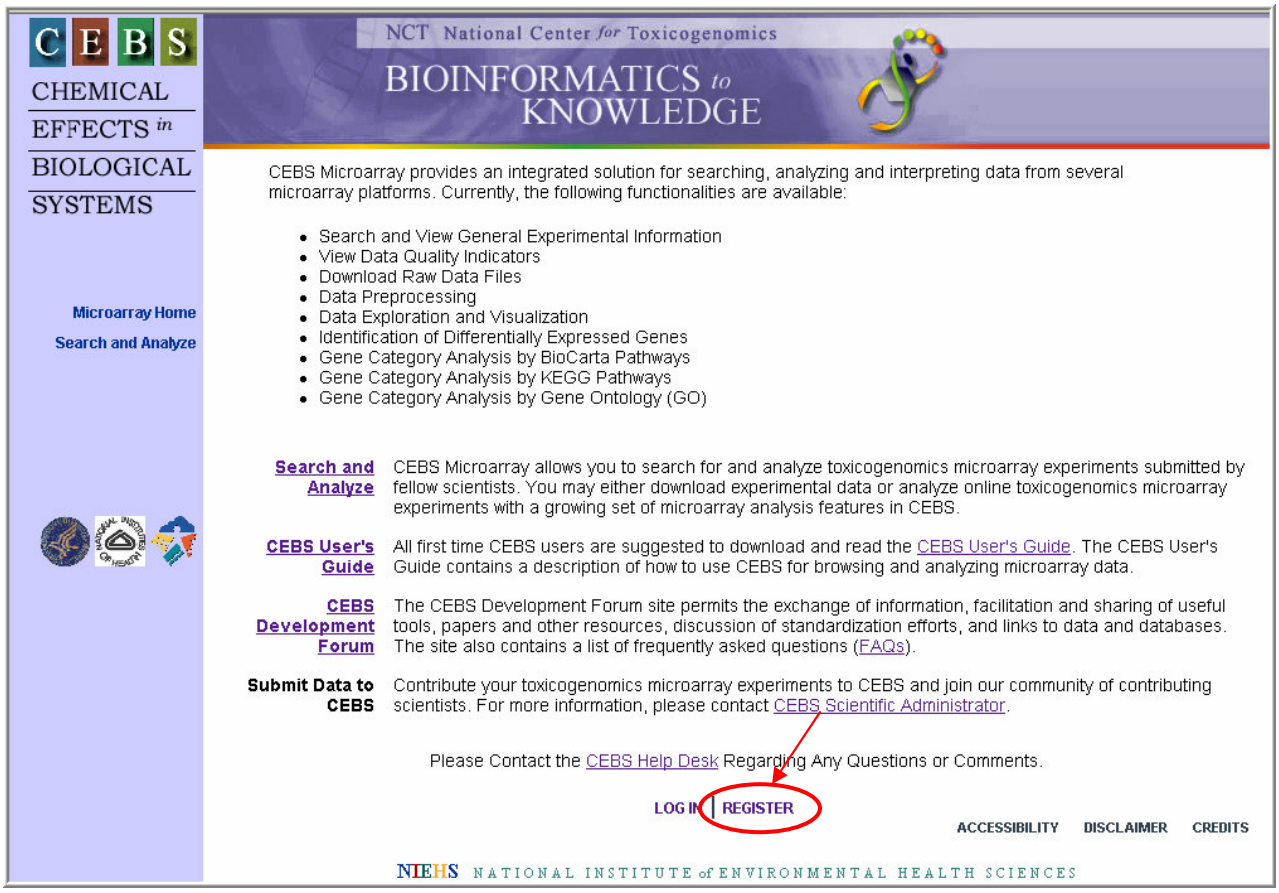


Figure 2 – Link to Account Registration Page

First, select the organization or laboratory to which you belong (Figure 3). If your organization is not listed, contact the CEBS Help Desk at cebsfeedback@list.niehs.nih.gov.

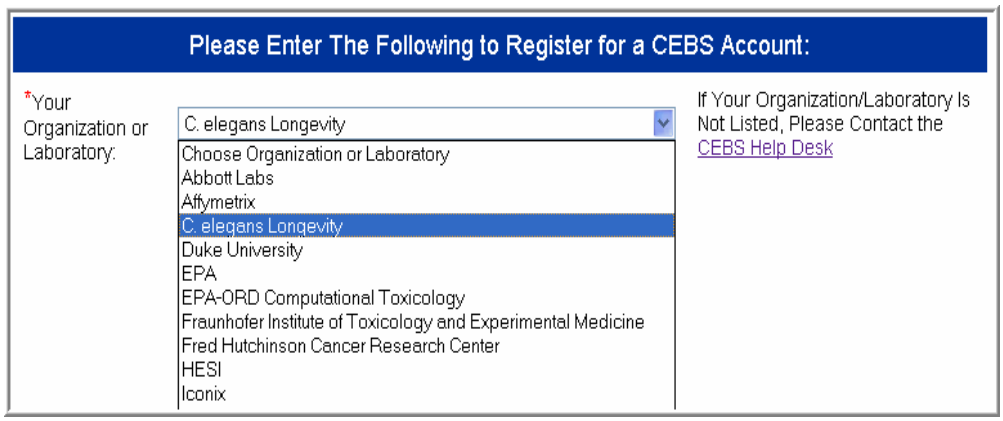


Figure 3 - Registration: Organization/Laboratory Selection

Once you have selected your organization/laboratory, the next page displayed will have a list of Principal Investigators for your organization or laboratory registered with CEBS.

If your Principal Investigator (PI) is registered in CEBS, fill out the information on the following page to register for a CEBS Microarray account (Figure 4). If your PI is not listed in CEBS, please contact the CEBS Help Desk at cebsfeedback@list.niehs.nih.gov

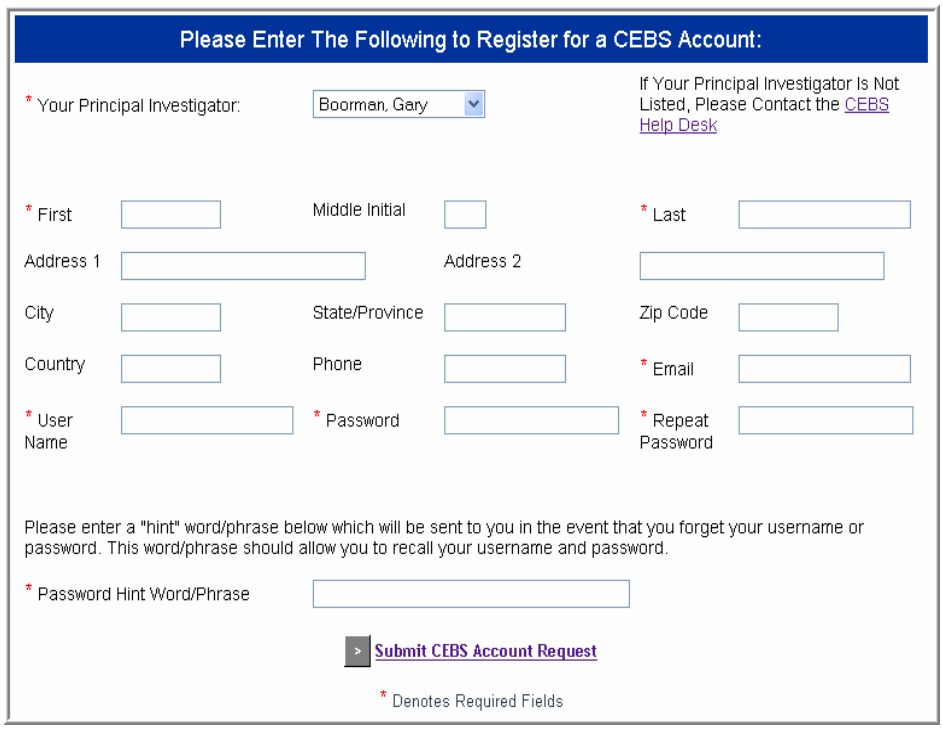


Figure 4 - CEBS Microarray Account Registration Page

- If your Principal Investigator is not in the drop-down box, have your PI send an email to cebsfeedback@list.niehs.nih.gov to be added to CEBS.
- When you get to the CEBS Microarray Account Registration page (Figure 4), enter as much information as possible. The items that have an "*" next to them are required fields.
- When you are done filling in the form, click on the **submit** button. This will send your information to the CEBS team. The CEBS team will verify your identity with the Principal Investigator you have selected. You will be sent an email notification once your information is verified and your account is activated.
- If you don't enter information into a required field, your login already exists or the passwords don't match, you will get an error message and will be prompted to re-enter the information. Note: user name and password are case sensitive.

Upon submission of your account request, the following confirmation page as shown in Figure 5 will be displayed. Again, once we've verified your membership in the PI's group you've selected, your account will be activated.

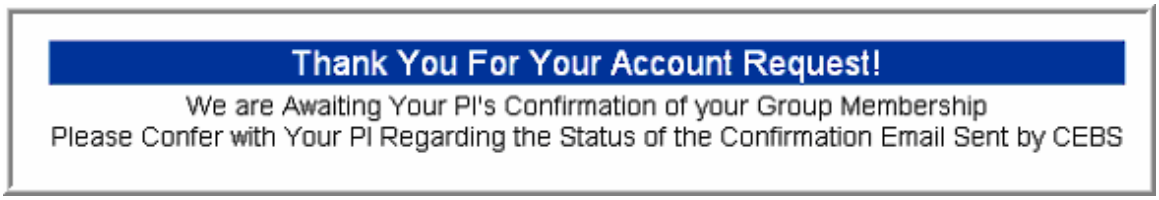


Figure 5 – CEBS Microarray Account Registration Confirmation Page

4. Logging In

In order to submit an experiment or access restricted-access data sets, you must have a user account to sign into the application. If you have already created a valid User Name and password, enter them into the Login page (Figure 6). Click the Login button to complete the login process.

Please note that the username and password fields are **case sensitive.

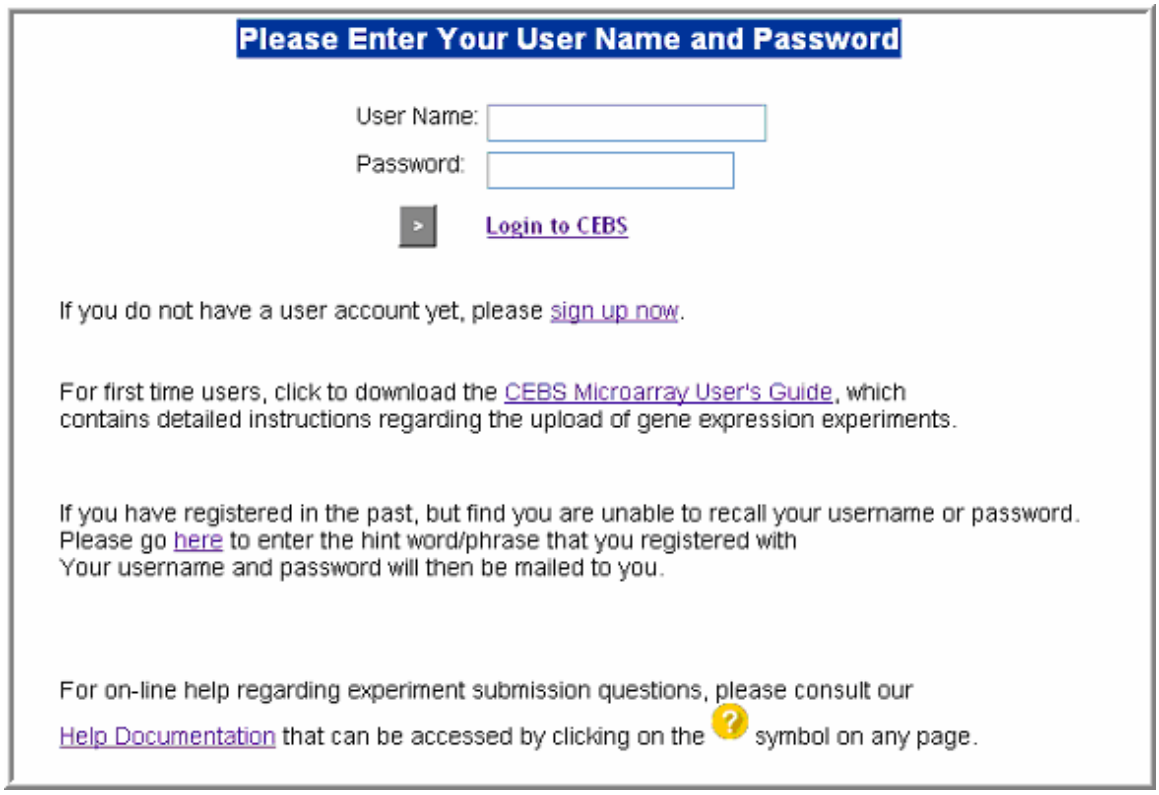


Figure 6 - CEBS Microarray Login Page

From the login page, you can register as a new user by clicking the “sign up now” link.

You can also download the user’s guide by clicking on the “CEBS Microarray User’s Guide” link.

If you have forgotten your username and/or password, select the “here” button to retrieve your account information (Figure 7). If you do not remember your hint word, please send an email to cebsfeedback@list.niehs.nih.gov.

4.1. Forgotten Password – Getting Password Hint

If you have forgotten your password, there is a link on the login page which will take you to the page displayed below.

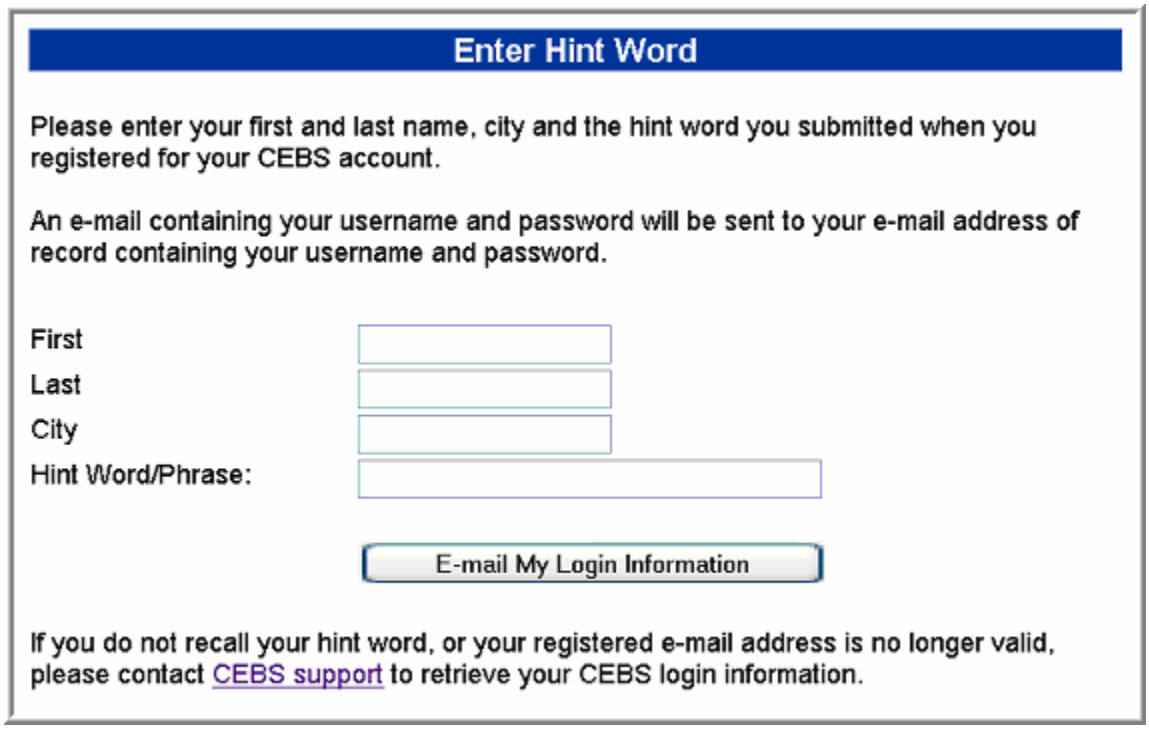


Figure 7 - Password Hint Page

Enter the required information, and your username and password will be emailed to the email address you initially registered. If you do not remember your hint word, or your email address is no longer valid, please send an email to cebsfeedback@list.niehs.nih.gov.

5. CEBS Microarray Home Page

The Microarray Home Page is the first page displayed when you enter the microarray portion of the CEBS website.

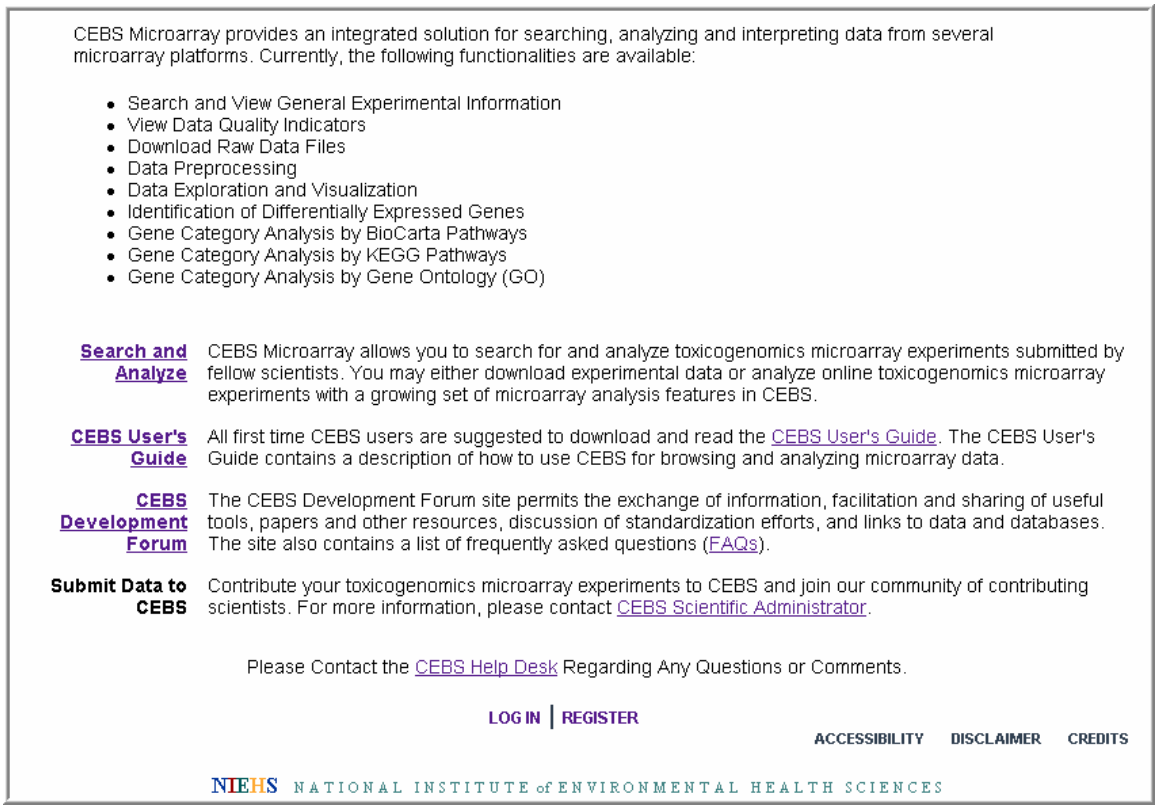


Figure 8 - Microarray Home Page Main Body

There are a number of link options on the Home page.

- **Search and Analyze** – This option will allow you to search for and analyze toxicogenomics microarray experiments submitted by fellow scientists. You also have the option to download the experimental data.
- **CEBS User's Guide** – This link will allow you to download and read the CEBS User's Guide. The users's guide contains a description of how to use CEBS for browsing and analyzing microarray data.
- **CEBS Development Forum** – This will take you to the CEBS Development Forum which permits the exchange of information, facilitation and sharing of useful tools, papers, and other resources, discussion of standardization efforts, and links to data and databases. This site also contains a list of FAQs.
- **Submit Data to CEBS** – To contribute data to CEBS, please contact the CEBS Scientific Administrator.

6. Browse Experiment Information

In order to search for an experiment that is not publicly accessible or visible to the public, you must login to the application. If you do not have a user account you must Register a New User Account. See section 3 of this document.

Here is a list of the different levels of experiment visibility:

- Private – visible only to users with the same Principal Investigator.
- Public – visible to anyone accessing the web site.
- TRC Consortium – visible to all users who's Principal Investigators are members of the Toxicogenomic Research Consortium (TRC).

To login to the system see Logging In, Section 4.

6.1. Experiment Search

On **CEBS Microarray Home Page (**Figure 1**)**, click on **Search and Analyze** on the left menu, or **Search and Analyze** within the body of the home page. You will enter the **Analysis Home Page** (Figure 9). This page summarizes the major functionalities supported by CEBS Analysis Tools. Currently CEBS provides Experiment Search, View, and Analysis as an integrated, streamlined solution. You will begin data viewing and / or analysis from the Experiment Search.

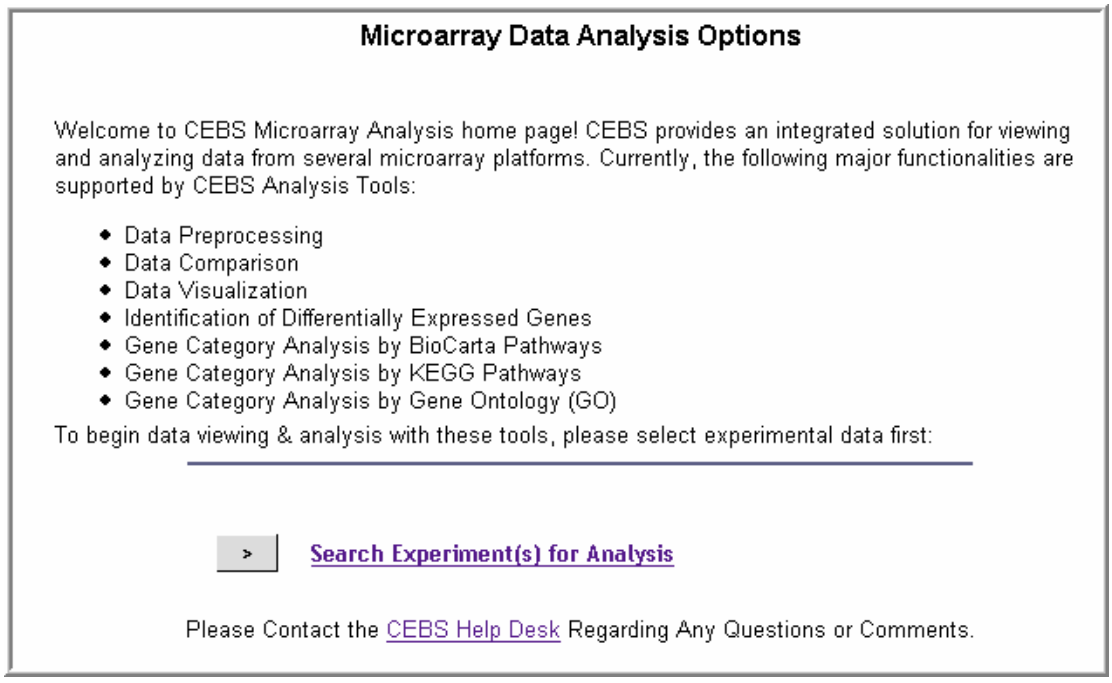


Figure 9 – CEBS Microarray Analysis Home Page

Click on **Search Experiment(s) for Analysis**, you will now enter **Experiment Search Page** (Figure 10).

If you have clicked **Search and Analyze** within the **CEBS Microarray Home Page**, you will omit the Analysis Home Page and directly enter **Experiment Search Page**.

You can search for experiments based on a combination of 6 fields. These fields are:

1. Experimental Id – text entry
2. Investigator's Name – drop-down list
3. Experiment Title – drop-down list
4. Tissue Name – drop-down list
5. Species – drop-down list
6. Image Processing Software – drop-down list

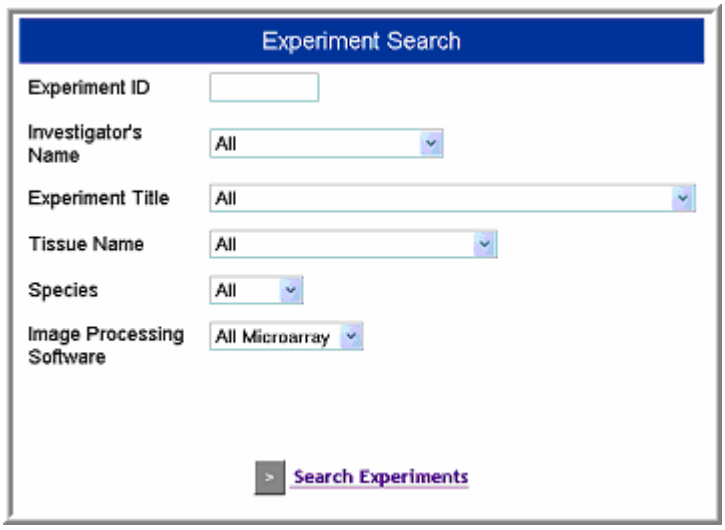


Figure 10 – Microarray Experiment Search Page

Once you have entered above search parameters, you can click on **Search Experiments** to begin the query.

6.2. Experiment Search Results

If there are no experiments matching your search criteria, the following screen will be displayed.

Figure 11 – Unsuccessful Search Results Page

If you are looking for an experiment recently submitted, remember that it must be fully curated before it can be accessed. You will receive an email when the experiment is curated and ready to be accessed.

Click on the Back button in the browser to return to the search page.

If there are experiments matching your search criteria, the following screen will be displayed (Figure 12).

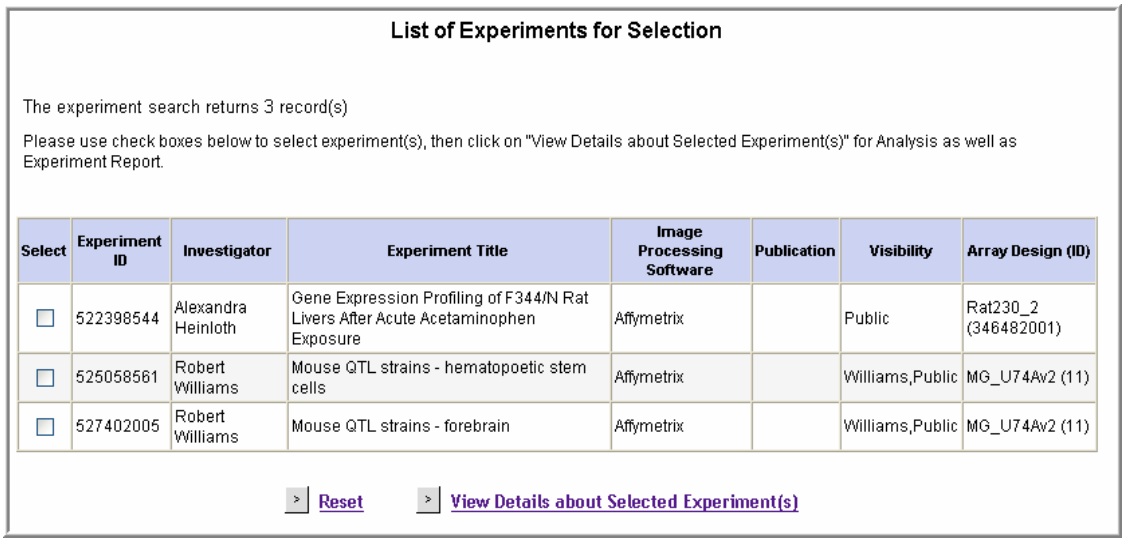


Figure 12 - Successful Search Results Page

The result screen (shown above) returns the following information about each experiment in the results list.

1. Experiment Id
2. Investigator Name
3. Experiment Title
4. Image Processing Software
5. Publication
6. Visibility
7. Array Design ID

To view the detailed information of the experiment(s) you are interested in, click on the checkbox in the **Select** column in front of the experiment(s) of

interest, then click on **View Details about Selected Experiment(s)**. Click on **Reset** to cancel the current selections and re-select.

Note: You may include ANY experiments that interest you at this time, regardless of their platforms or array designs. Subsequently, however, when you analyze multiple experiments together (Figure 14– Experiment Information Report Screen), you are only allowed to choose those experiments that have the same platform and matched array designs.

6.3. Experimental Details

Once you have clicked on **View Details about Selected Experiment(s)**, if one or more of the experiments you requested are of limited access, you will be prompted with the following message:

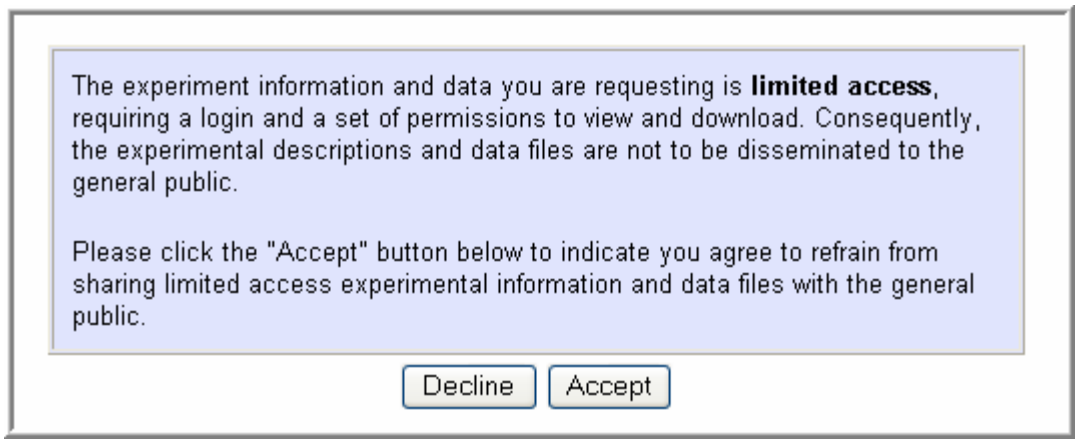
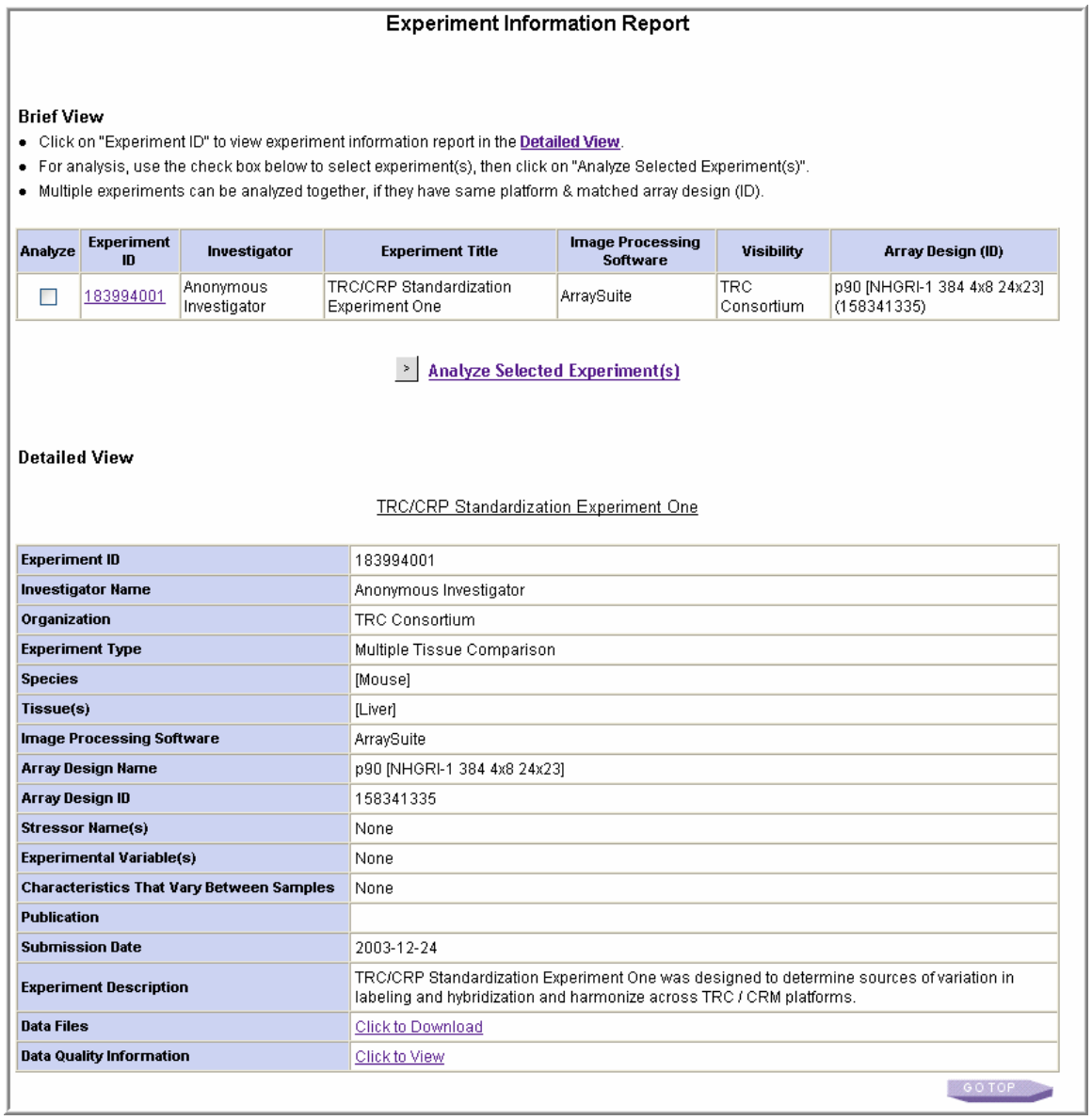


Figure 13 – Confidentiality Screen

You must choose **Accept** in order to continue to view more related experiment information. **Decline** returns to previous search screen. Once you have clicked on the **Accept** button, you will enter the **Experiment Information Report** page that contains more details about the experiment(s) you requested (Figure 14 – Experiment Information Report Screen).



Experiment Information Report

Brief View

- Click on "Experiment ID" to view experiment information report in the Detailed View.
- For analysis, use the check box below to select experiment(s), then click on "Analyze Selected Experiment(s)".
- Multiple experiments can be analyzed together, if they have same platform & matched array design (ID).

Analyze	Experiment ID	Investigator	Experiment Title	Image Processing Software	Visibility	Array Design (ID)
☐	183994001	Anonymous Investigator	TRC/CRP Standardization Experiment One	ArraySuite	TRC Consortium	p90 [NHGRI-1 384 4x8 24x23] (158341335)

Analyze Selected Experiment(s)

Detailed View

TRC/CRP Standardization Experiment One

Experiment ID	183994001
Investigator Name	Anonymous Investigator
Organization	TRC Consortium
Experiment Type	Multiple Tissue Comparison
Species	[Mouse]
Tissue(s)	[Liver]
Image Processing Software	ArraySuite
Array Design Name	p90 [NHGRI-1 384 4x8 24x23]
Array Design ID	158341335
Stressor Name(s)	None
Experimental Variable(s)	None
Characteristics That Vary Between Samples	None
Publication	
Submission Date	2003-12-24
Experiment Description	TRC/CRP Standardization Experiment One was designed to determine sources of variation in labeling and hybridization and harmonize across TRC / CRM platforms.
Data Files	Click to Download
Data Quality Information	Click to View

GO TOP

Figure 14 – Experiment Information Report Screen

This page consists of two sections: **Brief View** and **Detailed View**. The **Brief View** allows the user to review the experiment(s) they selected from the previous page. The **Detailed View** displays**,** for each experiment the user selected, more detailed information including Organization, Experiment Type, Species, Tissue(s), Stressor Name(s), Experimental Variable(s), Characteristics That Vary Between Samples, Submission Date, Experiment Description. Also provided are links for **Data Files Download** (see **Download Data Files** chapter 6) and **Data Quality Information** (see **Quality Control Information** chapter 8). Click on an **Experiment ID** link from **Brief View** brings down to the **Detailed View**, so that the user can view the detailed experiment report for this particular experiment. Click on the **Back to Top** link under each report to return to the **Brief View**.

6.4. Download Data Files

Once you have chosen an experiment and selected the link for data files download from the **Detailed View** of **Experiment Information Report** page, the **Date Files Download** page will be displayed, which contains links to experiment data files (Figure 15 - File Download Page).

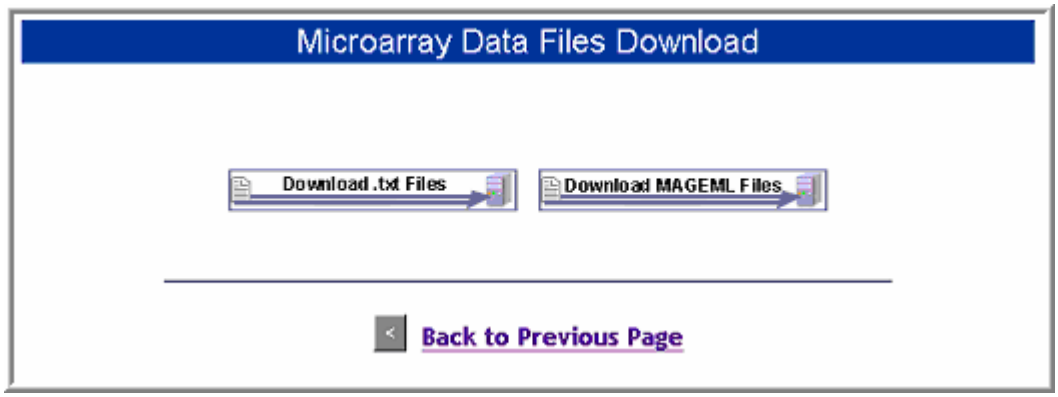


Figure 15 - File Download Page

In this example, the following files can be downloaded:

- .txt files
- MAGEML file

Other data files that may also be available for download include:

- .cel files
- Sample files
- .rpt files

Once a Download File button has been clicked, the **File Download** dialog box will appear (**Figure 16– File Download Dialog Box**)

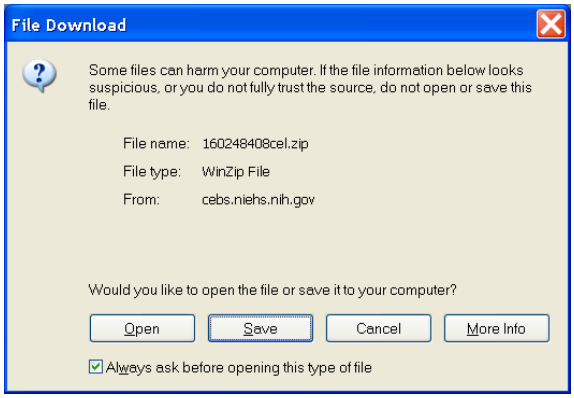


Figure 16– File Download Dialog Box

Click **Save** and then choose the location where you wish to save the file and click **Save** again (**Figure 17 – File Download Save Options**).

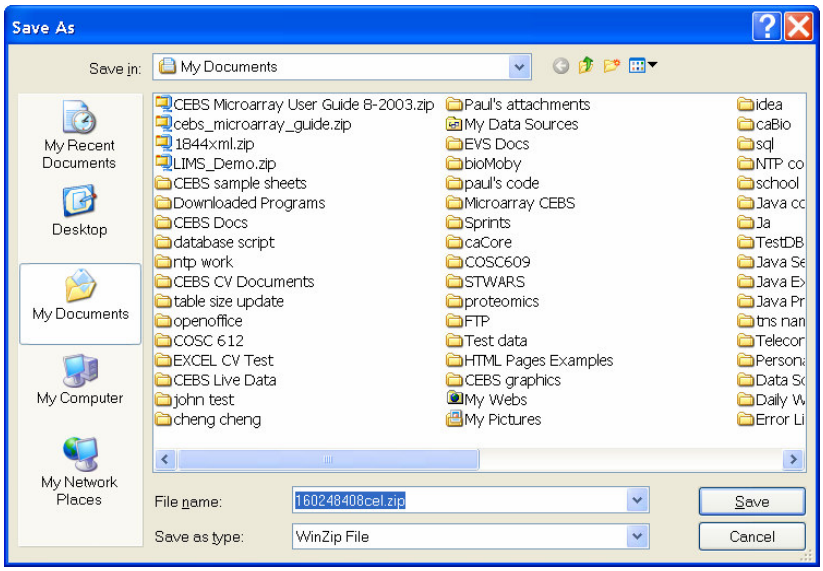


Figure 17 – File Download Save Options

The files downloaded will be archived with WinZip and will have the .zip file extension. You must first use WinZip, or some other similar file compression program, to unzip the file before you can access the data.

7. Data Analysis

CEBS provides a set of tools for analyzing data from several microarray platforms. A user can retrieve data from all or a subset of arrays in an experiment and perform analysis. The analysis workflow starts with data preprocessing. CEBS offers a variety of options to preprocess the data, curtailed to the platform and/or image processing software. The processed data set can then be visualized in exploratory plots and is stored at the server for subsequent analysis. Currently, CEBS has implemented a data comparison tool to allow for identification of differentially expressed genes by magnitude of difference in expression, and/or statistical significance of the difference, based on user specification. The results from the comparison tool can then be utilized by CEBS tools for biological analysis and visualization, which facilitate identification of the categories of biological activities (as provided by BioCarta, KEGG and Gene Ontology) that demonstrate the most significant difference in gene expression.

7.1. Select Experiment(s) to Analyze:

The users can select one or multiple experiments that they wish to analyze from the **Brief View** of **Experiment Information Report** page (Figure 14 – Experiment Information Report Screen). Click on **Analyze Selected Experiment(s)** to invoke **Select Arrays for Analysis** pages that allow the users to view and select arrays as input, in each of the experiments of interest (see **Data Selection and Retrieval** chapter for details).

Note: multiple experiments can be analyzed together ONLY when they have the same platform and matched array design ID.

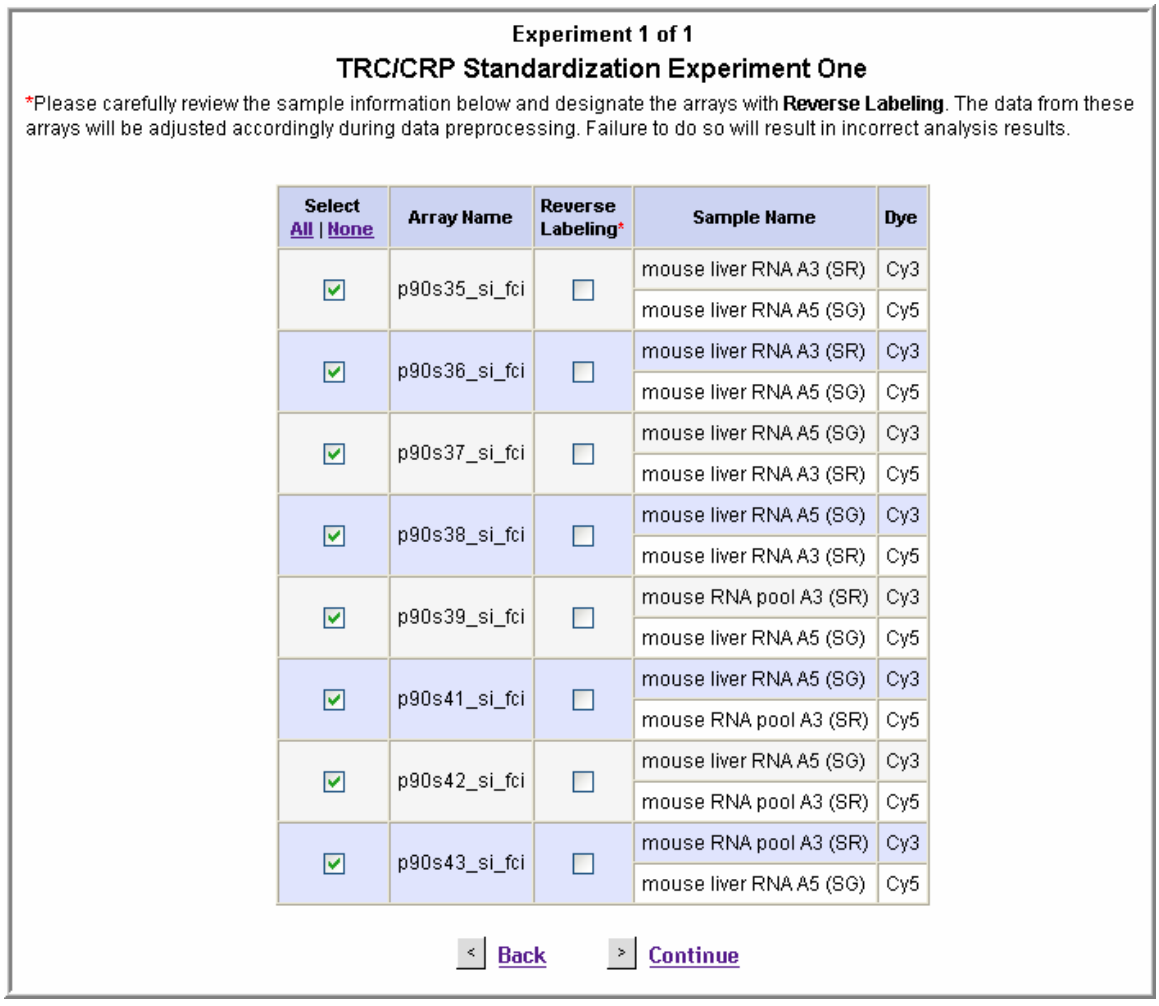
7.2. Data Selection and Retrieval:

Once you have clicked on **Analyze Selected Experiment(s**) from **Experiment Information Report** page (Figure 14 – Experiment Information Report Screen), you will enter **Select Arrays for Analysis** page (Figure 18). The page presents a complete list of arrays in the selected experiment, together with information that is captured for each array.

The user can select all the arrays, or a subset of them, in an experiment, as input for subsequent analysis. Click on **Select All** or **None** to select all arrays or de-select them all. For two-channel microarray data, the user can indicate the reverse labeled (dye swapping) arrays, so the two channels will be reversed back during data processing. The only exception is the output of ArraySuite software,

from which the output is already based on the sample and reference selection instead of the type of dye labeling. In this case, the data from the two channels will not be reversed during the preprocessing.

If multiple experiments are selected from **Experiment Information Report** page (plus, they have the same platform and matched array designs), the array information for all the experiments will be displayed, with one experiment on each page, in the order of the experiments displayed. Click on **Continue** to navigate through, view and select arrays in every experiment, until you have made array selections for all the experiments you have chosen for analysis. Finally, click on **Continue** to complete the data set selection and retrieval. Now you are ready to perform data set preprocessing (see **Preprocessing** chapter 7).



Experiment 1 of 1

TRC/CRP Standardization Experiment One

*Please carefully review the sample information below and designate the arrays with **Reverse Labeling**. The data from these arrays will be adjusted accordingly during data preprocessing. Failure to do so will result in incorrect analysis results.

Select All | None	Array Name	Reverse Labeling*	Sample Name	Dye
☑	p90s35_si_fci	☐	mouse liver RNA A3 (SR)	Cy3
			mouse liver RNA A5 (SG)	Cy5
☑	p90s36_si_fci	☐	mouse liver RNA A3 (SR)	Cy3
			mouse liver RNA A5 (SG)	Cy5
☑	p90s37_si_fci	☐	mouse liver RNA A5 (SG)	Cy3
			mouse liver RNA A3 (SR)	Cy5
☑	p90s38_si_fci	☐	mouse liver RNA A5 (SG)	Cy3
			mouse liver RNA A3 (SR)	Cy5
☑	p90s39_si_fci	☐	mouse RNA pool A3 (SR)	Cy3
			mouse liver RNA A5 (SG)	Cy5
☑	p90s41_si_fci	☐	mouse liver RNA A5 (SG)	Cy3
			mouse RNA pool A3 (SR)	Cy5
☑	p90s42_si_fci	☐	mouse liver RNA A5 (SG)	Cy3
			mouse RNA pool A3 (SR)	Cy5
☑	p90s43_si_fci	☐	mouse RNA pool A3 (SR)	Cy3
			mouse liver RNA A5 (SG)	Cy5

< Back > Continue

Figure 18 – Select Arrays for Analysis

7.3. Preprocessing:

CEBS offers a variety of options for data filtering, transformation and normalization for both one-channel and two-channel microarray data. The goal is to allow for meaningful cross-array comparison of gene expression data.

7.3.1. Two-channel arrays:

CEBS supports preprocessing of two-channel microarray data from several types of image processing software, including Affymetrix (Figure 19), GenePix (Figure 20), ArraySuite (Figure 21), Agilent's Feature Extraction software (Figure 22), and Generic One Channel (Figure 23). For GenePix, the raw spot intensity values will be used for preprocessing. In the cases of Agilent and ArraySuite, the users can either use the derived data (e.g. log-ratio, or ratio) that is computed by the image processing software, or use the raw spot intensity values.

The raw intensity data from two-channel arrays is filtered by an intensity threshold, and/or flags and other spot-level quality indicators from each platform based on user-specifications. The user has the option to perform **Linear Normalization**, or within-array normalization using **LOWESS** on log-ratio vs. log-product of intensities from both channels, using a user-specified **Span** value. A **Cross-array Normalization** algorithm based on inter-quarter range or median-absolute deviation is provided.

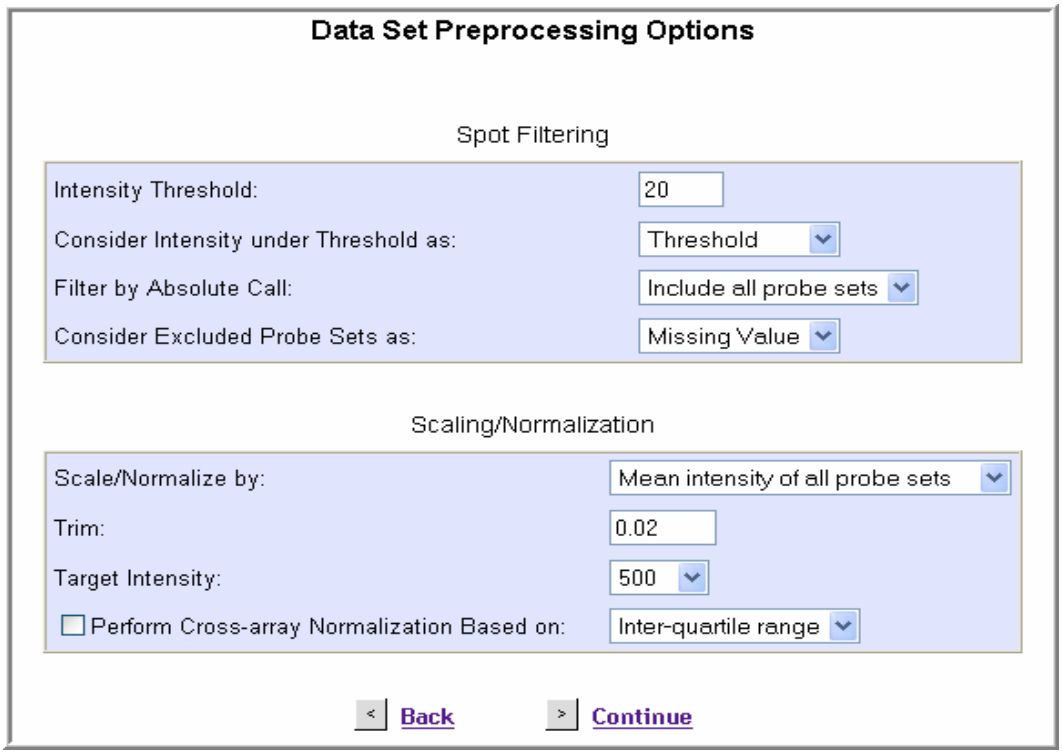


Figure 19 –Affymetrix preprocessing page

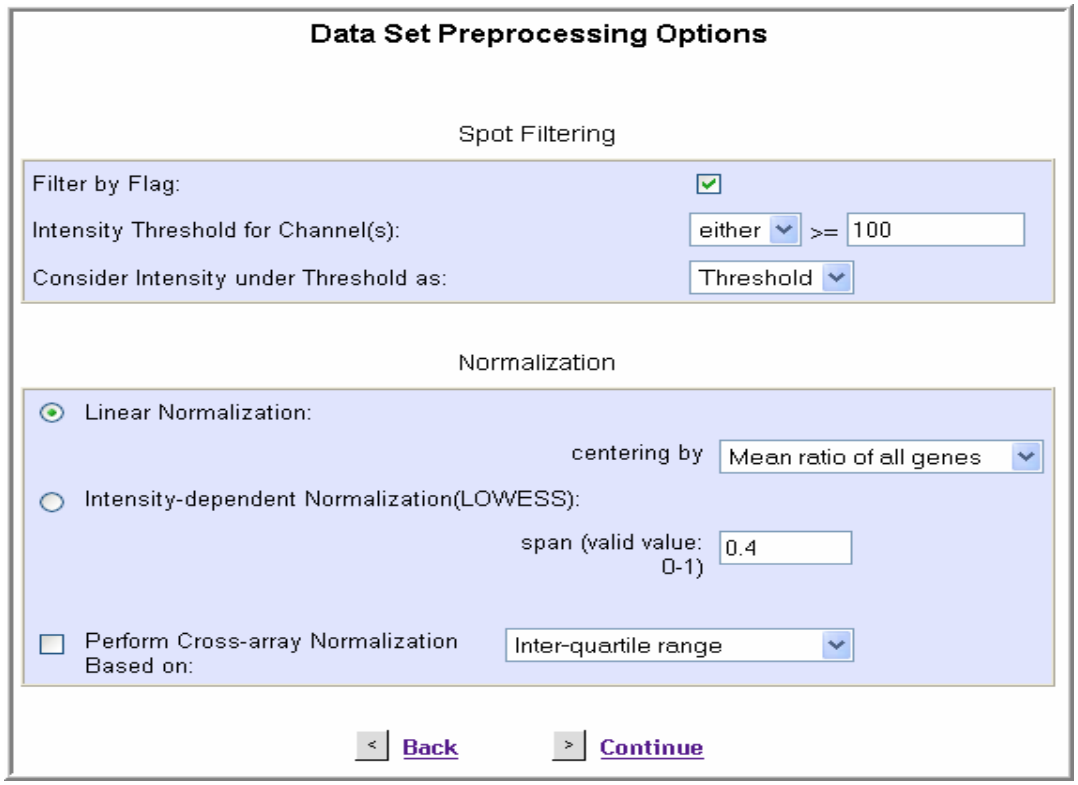


Figure 20 - GenePix preprocessing page

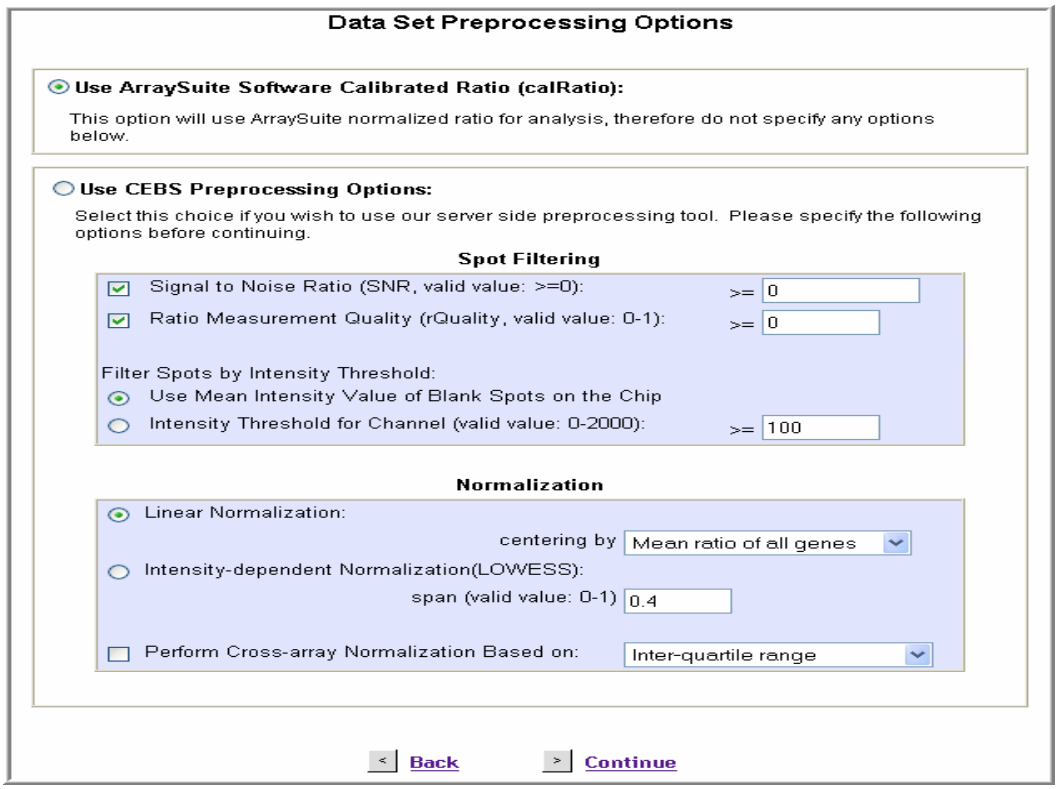


Figure 21 –ArraySuite Preprocessing Page

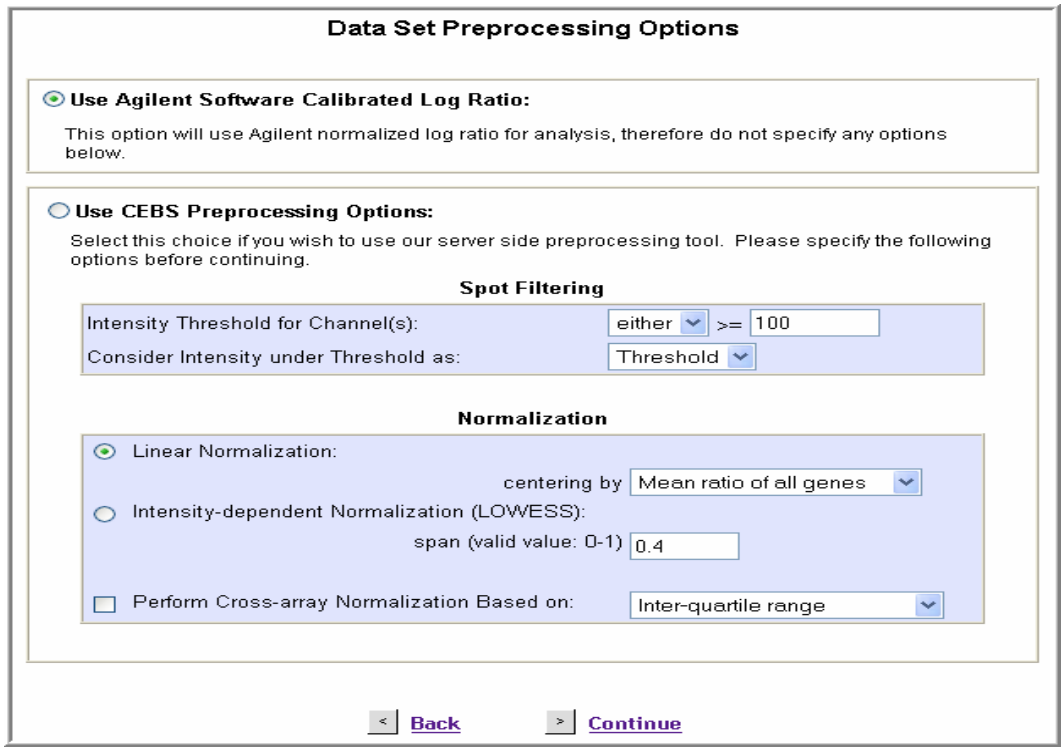


Figure 22 - Agilent preprocessing Page

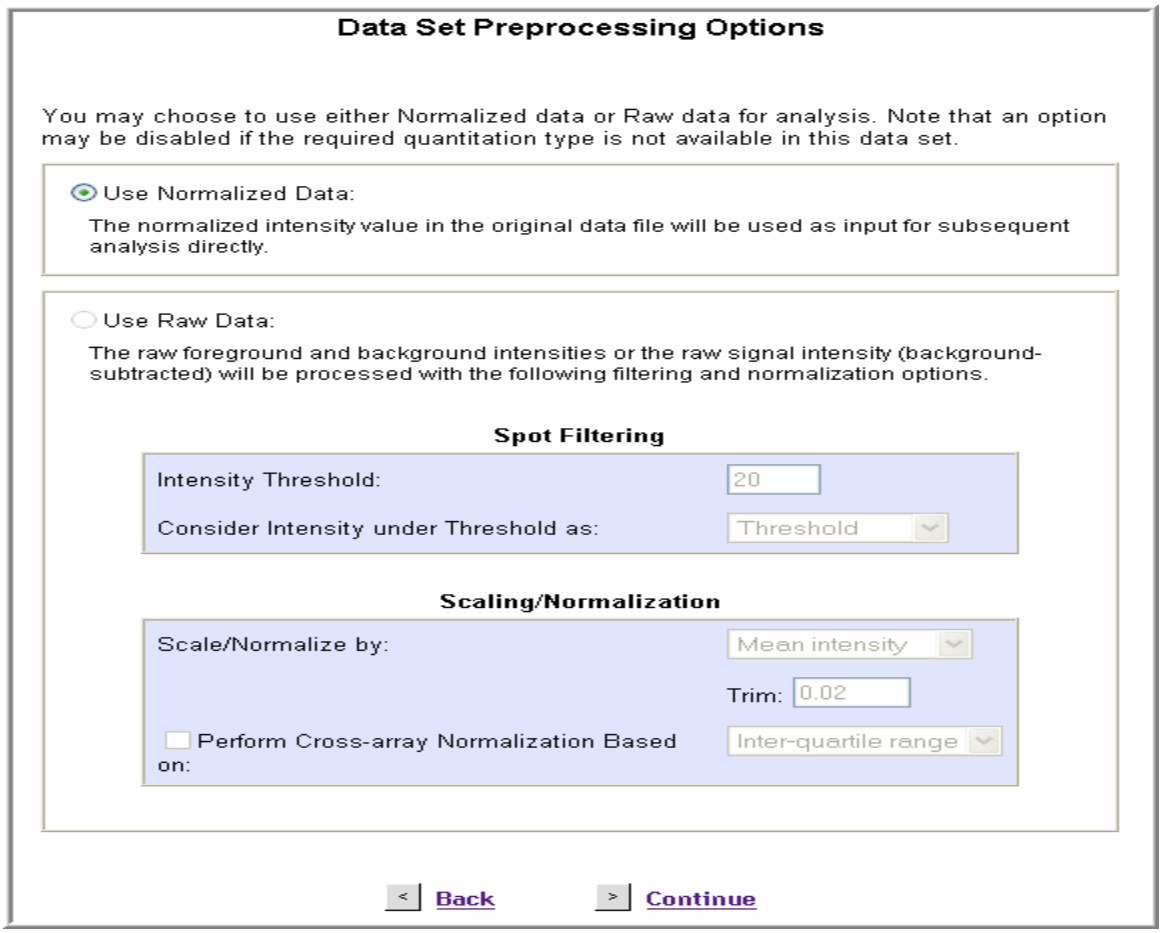


Figure 23 - Generic One Channel preprocessing page

7.3.2. One-channel arrays

The Affymetrix average difference or signal data is filtered based on the absolute call or detection call, as the user specified. The intensity or signal values are then scaled so that the mean or median intensity of each array is to be a user-specified target intensity. A small percentage of extreme values (as specified by **Trim**) are excluded when calculating means for scaling. The intensity values under a user-specified threshold are then adjusted to the threshold or as missing values for future analysis (Figure 19). A **Cross-array Normalization** algorithm based on inter-quartile range is provided.

7.4. Visualization of Microarray Gene Expression Data

After the data set preprocessing, CEBS provides the users with exploratory plots, an intuitive way to obtain gene expression data and to judge the comparability of the data sets that are selected from single or multiple experiments. The visual presentation of data can serve as part of an explorative analysis to help users determine the general trends in the data. Based on the visual presentation of data, the users can also determine if different options or parameters for normalization should be used.

On the options page (Figure 24), click on **Visualize Preprocessed Data** to view the graphic representation of preprocessed data displayed in the next page (Figure 25-Figure 27). Alternatively, you may choose to begin microarray data analysis by clicking on **Analyze Preprocessed Data**.

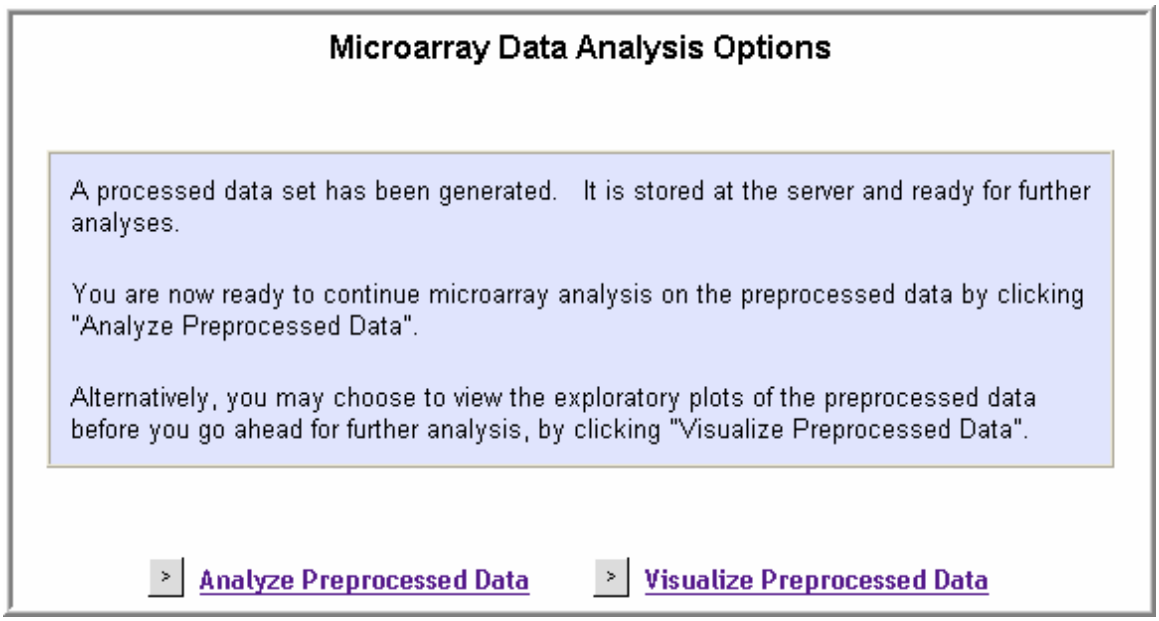


Figure 24 – Analysis Options

Currently, CEBS provides three types of plots for the purpose of visualization: Box-Whisker Plot, Dendrogram, and Multi-dimentional Scaling Plot.

The **Box-Whisker Plot (**Figure 25**)** portrays the essential data distribution characteristics, such as maximum, minimum, median, lower quartile, upper quartile, and outliers, for each of the arrays the user selected for analysis (see **Data Selection and Retrieval** chapter 7.2).

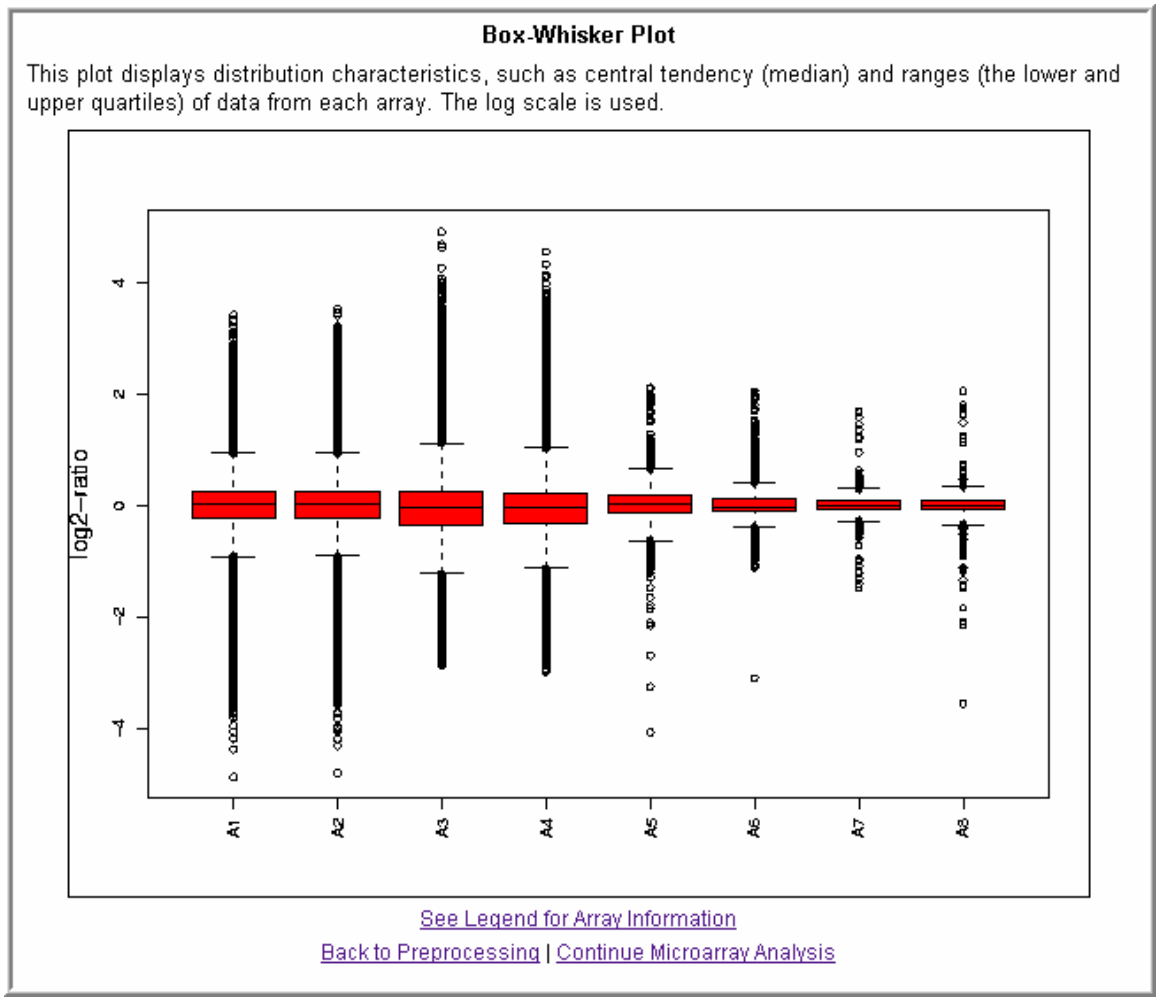


Figure 25 – Box-Whisker Plot – Microarray Visualization Tool

The **Dendrogram** (Figure 26) allows the users to view the relative similarities between arrays and hierarchical clustering of arrays, by clustering the arrays with similar patterns. The distances used for clustering are 1-pearson correlations between expression value (in log scale) from arrays. The left axis shows the distances between arrays or average distances between clusters of arrays.

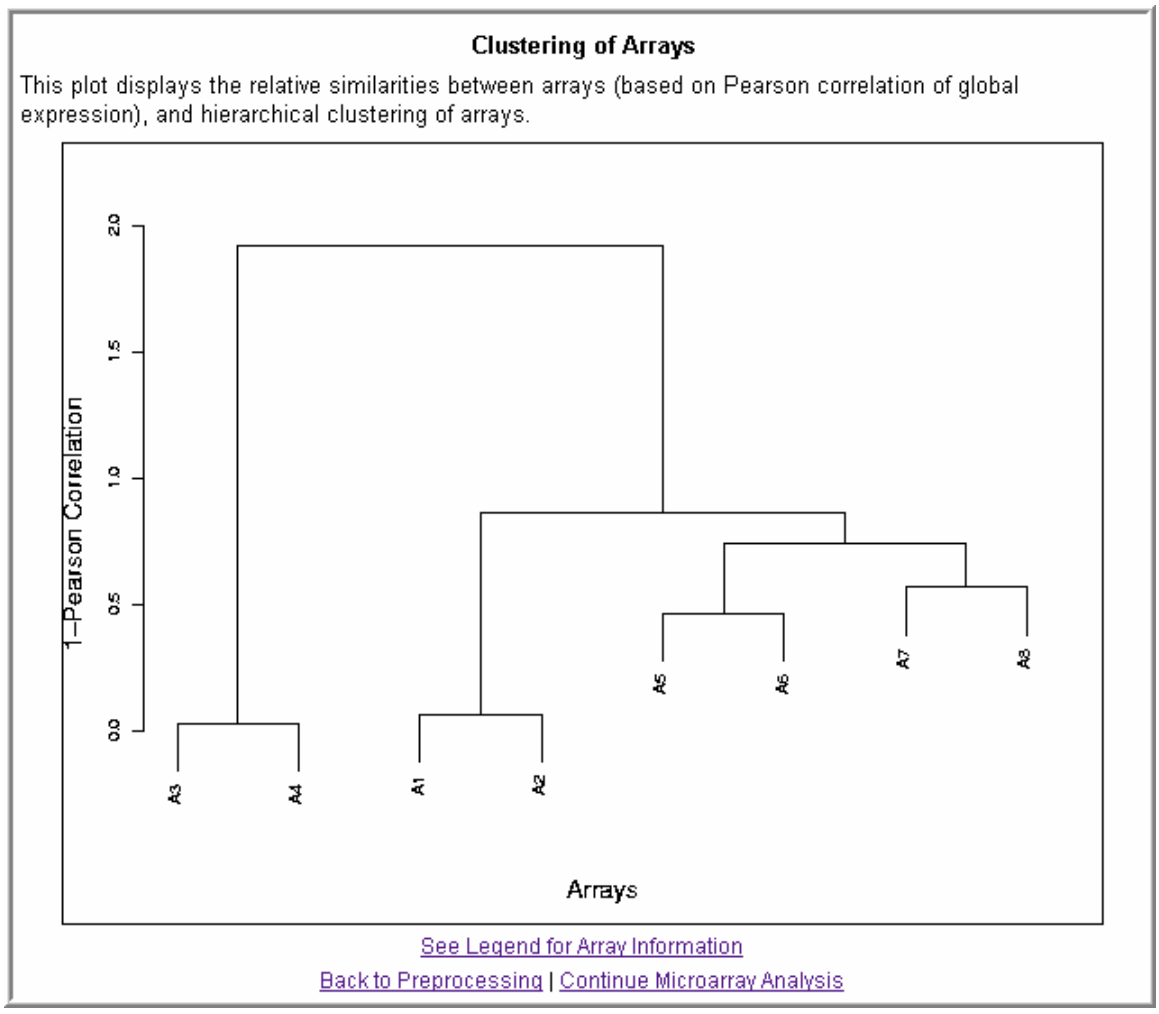


Figure 26 – Dendrogram – Microarray Visualization Tool

The **Multi-dimentional Scaling Plot** (MDS Plot, Figure 27) projects the relative distances between arrays to two-dimentional space with a multi-dimentional scaling algorithm and portrays the relative similarities between arrays, i.e. arrays with expression similarities cluster close to one another.

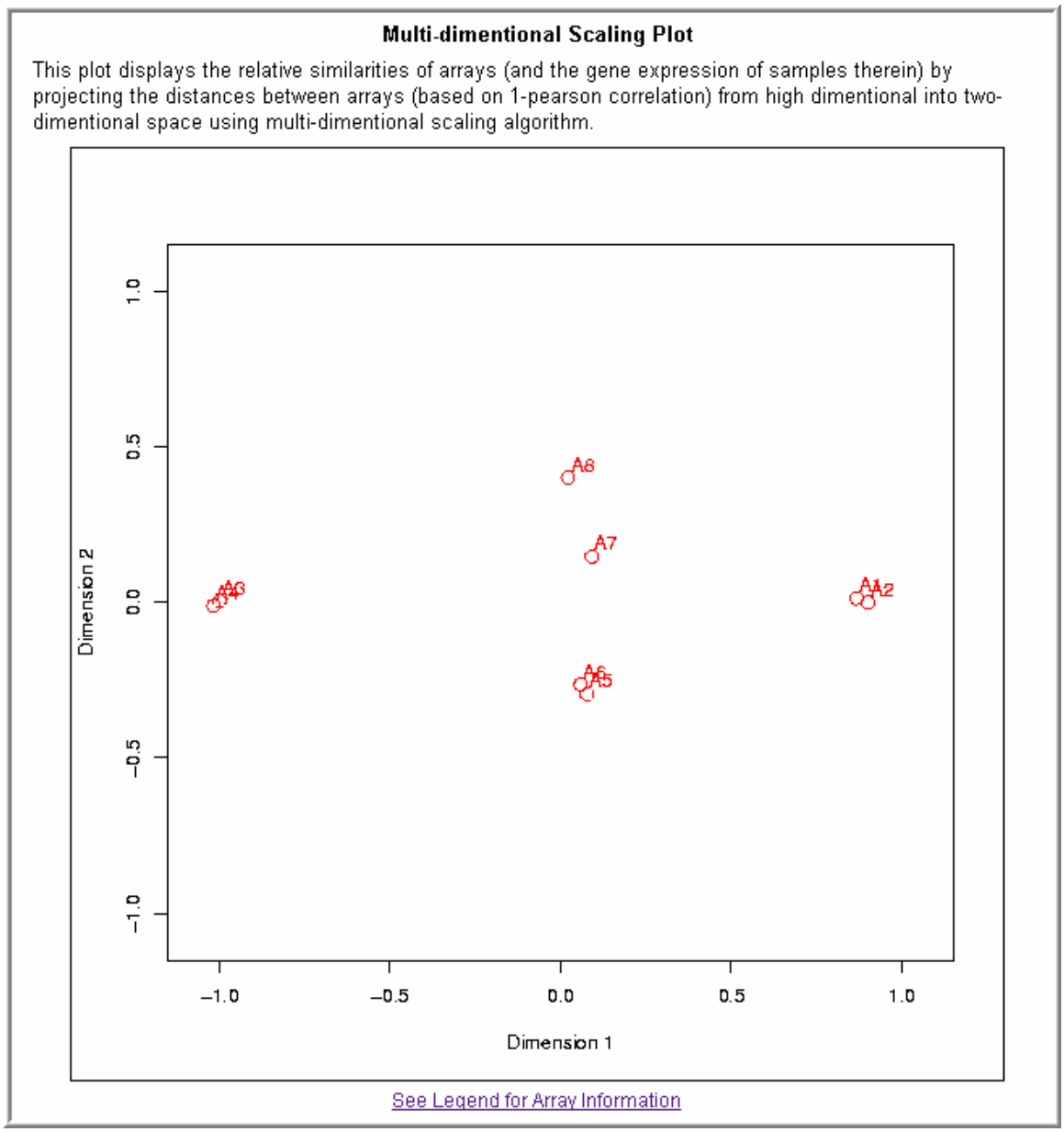
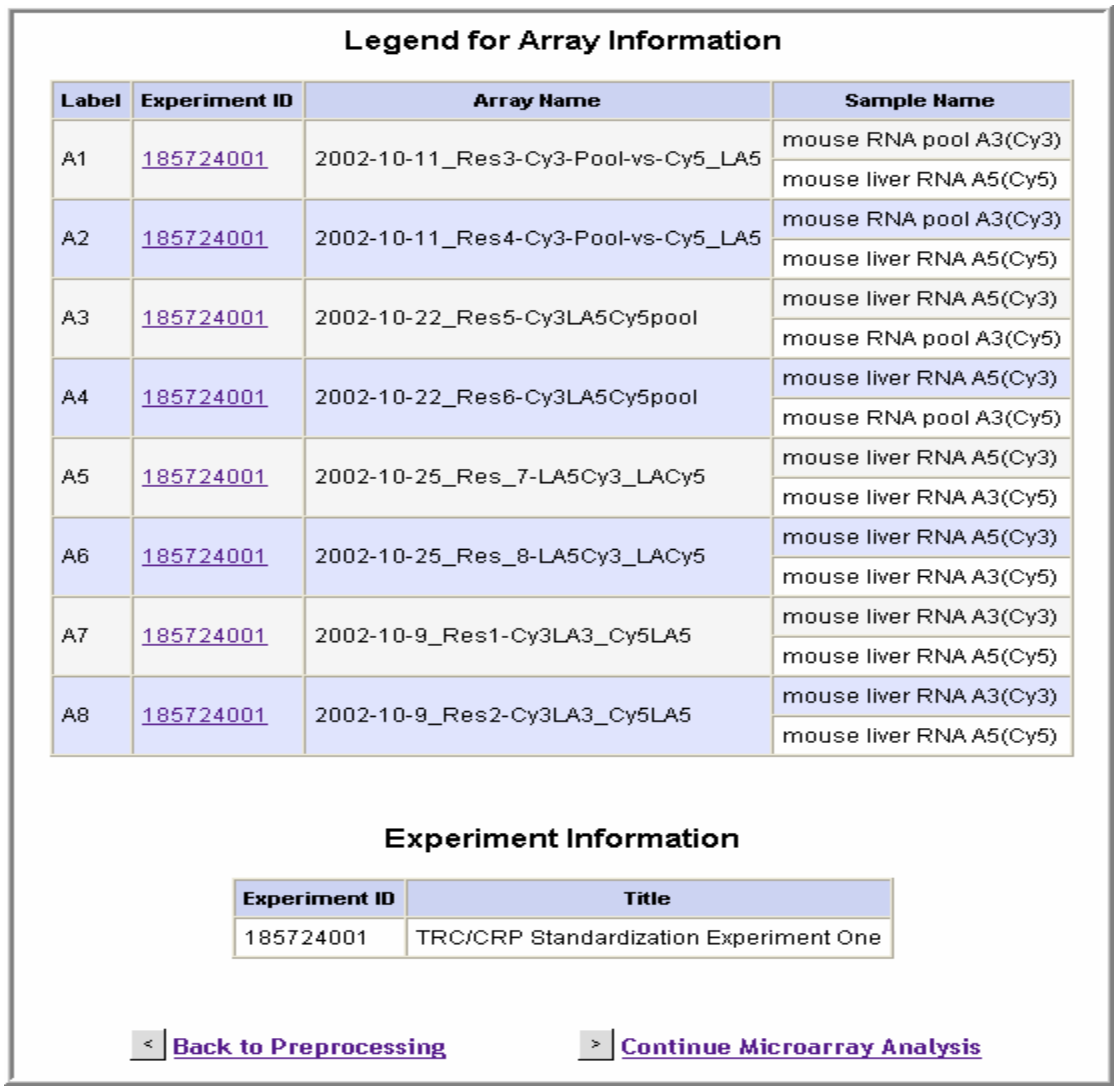


Figure 27 – Multi-dimentional Scaling Plot – Microarray Visualization Tool

To construct above exploratory plots in a convenient and intuitive way, a series of color codes are used to represent samples from different experiments. The link **See Legend for Array Information** under each plot image directs the user to detailed interpretation of the labels (such as A1, A2, B3, etc.) in the plots

corresponding to the arrays used in preprocessing. Information such as Experiment ID, Array Name and Sample Name is provided. In addition, Experiment Title is also displayed for each Experiment ID of the array samples in preprocessing (Figure 28).



Legend for Array Information

Label	Experiment ID	Array Name	Sample Name
A1	185724001	2002-10-11_Res3-Cy3-Pool-vs-Cy5_LA5	mouse RNA pool A3(Cy3)
			mouse liver RNA A5(Cy5)
A2	185724001	2002-10-11_Res4-Cy3-Pool-vs-Cy5_LA5	mouse RNA pool A3(Cy3)
			mouse liver RNA A5(Cy5)
A3	185724001	2002-10-22_Res5-Cy3LA5Cy5pool	mouse liver RNA A5(Cy3)
			mouse RNA pool A3(Cy5)
A4	185724001	2002-10-22_Res6-Cy3LA5Cy5pool	mouse liver RNA A5(Cy3)
			mouse RNA pool A3(Cy5)
A5	185724001	2002-10-25_Res_7-LA5Cy3_LACy5	mouse liver RNA A5(Cy3)
			mouse liver RNA A3(Cy5)
A6	185724001	2002-10-25_Res_8-LA5Cy3_LACy5	mouse liver RNA A5(Cy3)
			mouse liver RNA A3(Cy5)
A7	185724001	2002-10-9_Res1-Cy3LA3_Cy5LA5	mouse liver RNA A3(Cy3)
			mouse liver RNA A5(Cy5)
A8	185724001	2002-10-9_Res2-Cy3LA3_Cy5LA5	mouse liver RNA A3(Cy3)
			mouse liver RNA A5(Cy5)

Experiment Information

Experiment ID	Title
185724001	TRC/CRP Standardization Experiment One

< Back to Preprocessing > Continue Microarray Analysis

Figure 28 – Legend for the Microarray Visualization Tools

Click on **Continue Microarray Analysis** to go to the sample grouping/comparison page to begin analysis (Figure 29). Alternatively, the users may choose **Back to Preprocessing** to re-do preprocessing with a different set of methods and parameters.

7.5. Comparison Analysis and Identification of Differentially-expressed Genes

CEBS provides server-side tools to compare samples or groups of samples to identify differentially expressed genes, by the magnitude of the difference (e.g. fold change) in expression levels and/or statistical significance of the differences.

Currently, CEBS provides two types of comparison (Figure 29): (1) **Comparison of Two Groups of Arrays**: the user can compare arrays from two biological conditions, designated A and B, with or without replicates. This type of comparison applies to Affymetrix and two-channel array data with common reference. (2) **Comparison of Samples in the Same Arrays**: the user can compare samples from different channels in the same arrays. This type of comparison applies to two-channel array data only.

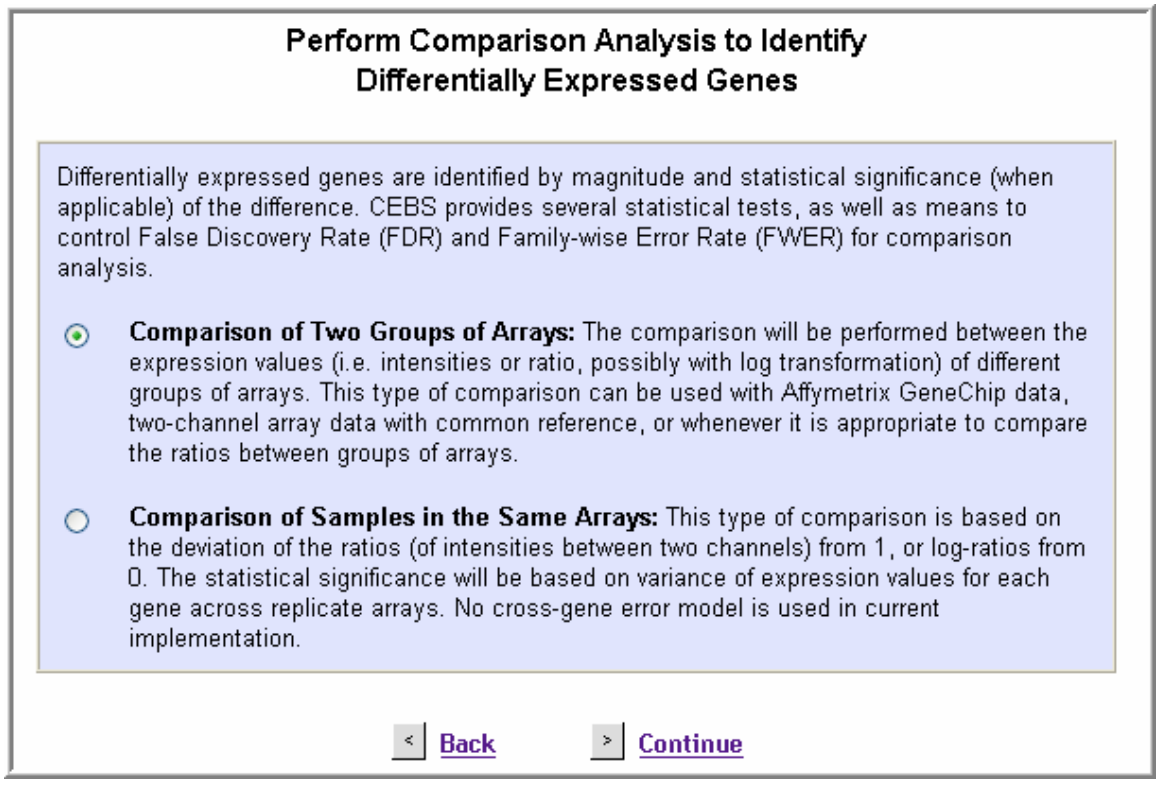
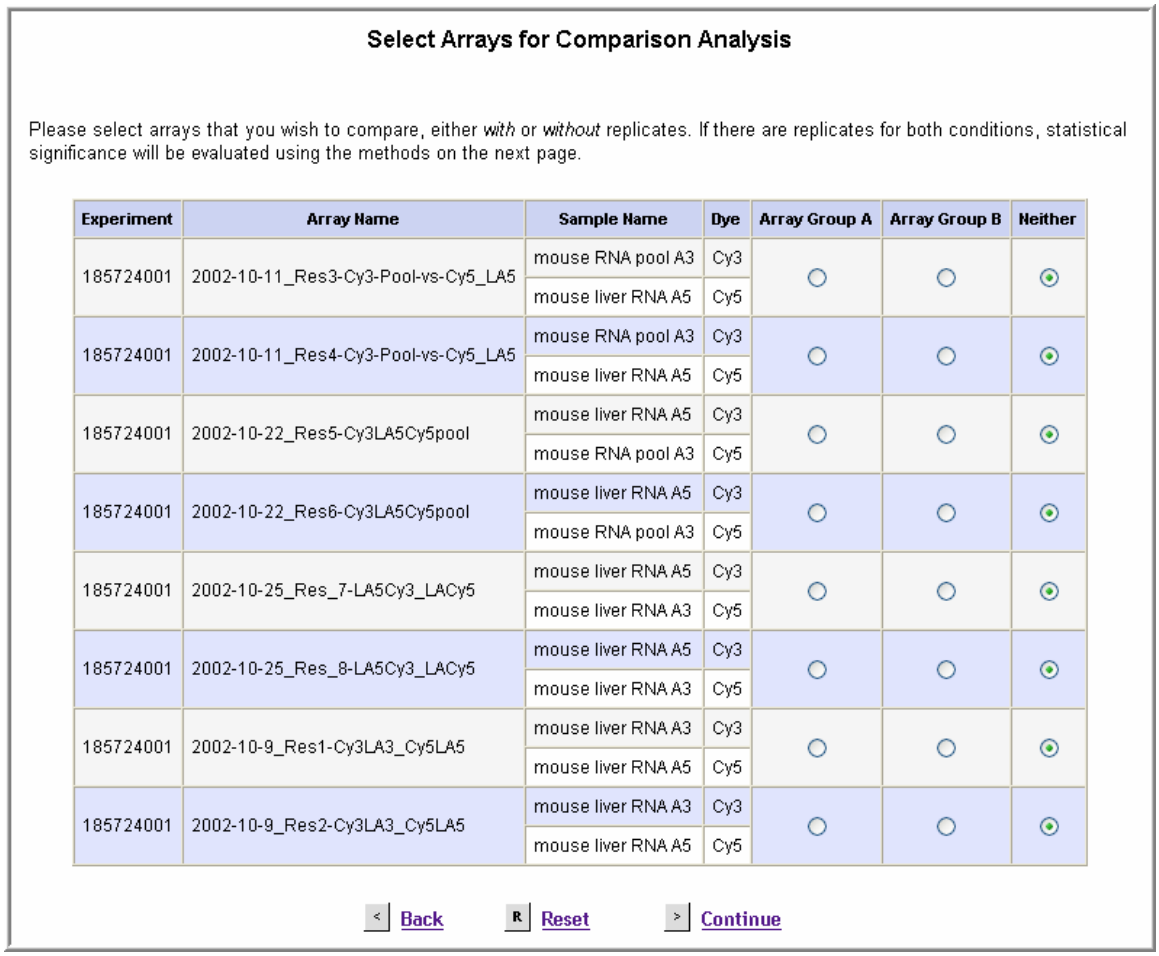


Figure 29 – Comparison Analysis Page

7.5.1. Assign samples/arrays for comparison

Type I: Comparison of Two Groups of Arrays

Choose this type of comparison and click on **Continue**, you will enter **Select Arrays for Comparison Analysis** page that allows you to select arrays that you wish to compare and assign them to condition **A** or **B** (Figure 30). Note that you may make **single-sample comparison** by comparing single array of A against another single array of B, or **multiple-sample comparison** by comparing multiple arrays of A against multiple arrays of B. Click on **Reset** to cancel the current selections and re-select. Click on **Continue** to come to the next page where you are provided with a combination of comparison methods and parameters (Figure 32).



Select Arrays for Comparison Analysis

Please select arrays that you wish to compare, either *with* or *without* replicates. If there are replicates for both conditions, statistical significance will be evaluated using the methods on the next page.

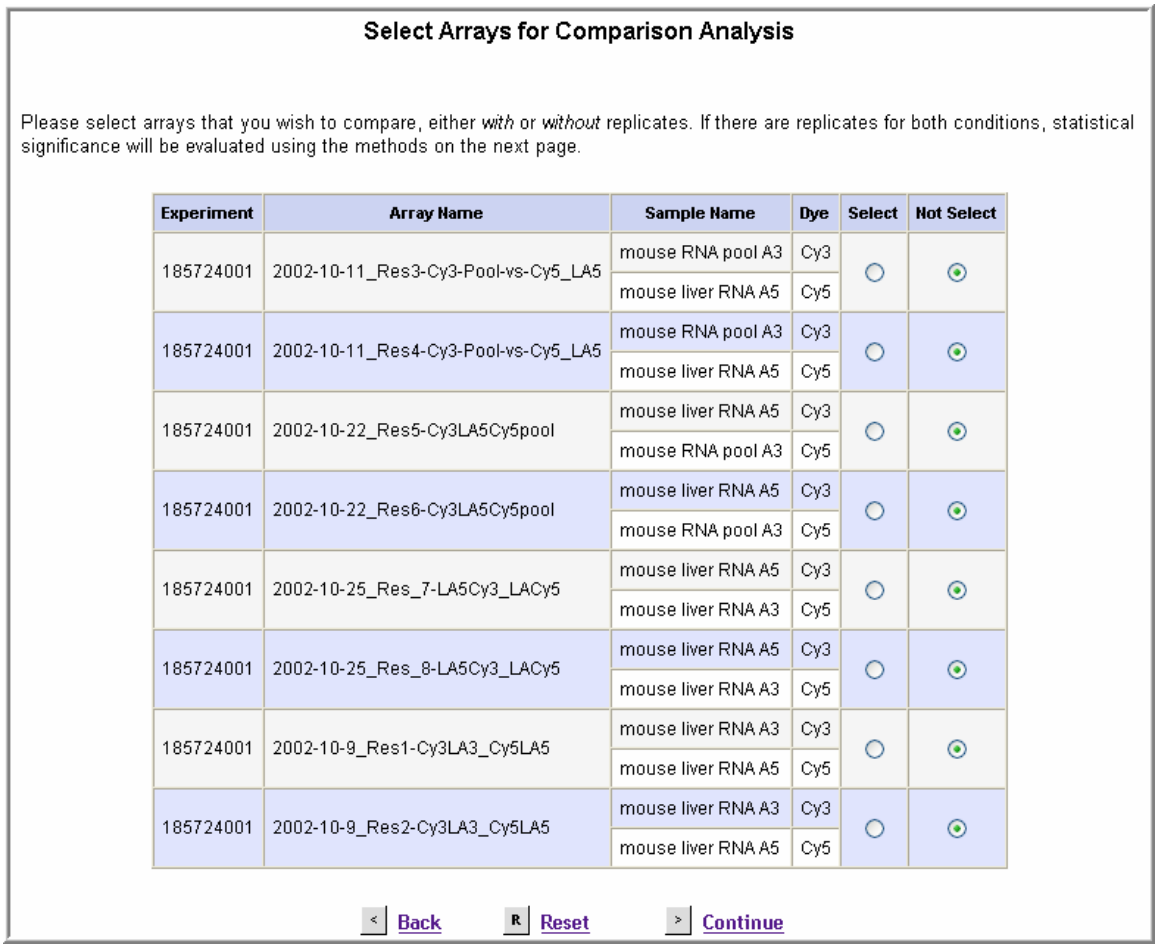
Experiment	Array Name	Sample Name	Dye	Array Group A	Array Group B	Neither
185724001	2002-10-11_Res3-Cy3-Pool-vs-Cy5_LA5	mouse RNA pool A3	Cy3	○	○	◉
		mouse liver RNA A5	Cy5			
185724001	2002-10-11_Res4-Cy3-Pool-vs-Cy5_LA5	mouse RNA pool A3	Cy3	○	○	◉
		mouse liver RNA A5	Cy5			
185724001	2002-10-22_Res5-Cy3LA5Cy5pool	mouse liver RNA A5	Cy3	○	○	◉
		mouse RNA pool A3	Cy5			
185724001	2002-10-22_Res6-Cy3LA5Cy5pool	mouse liver RNA A5	Cy3	○	○	◉
		mouse RNA pool A3	Cy5			
185724001	2002-10-25_Res_7-LA5Cy3_LACy5	mouse liver RNA A5	Cy3	○	○	◉
		mouse liver RNA A3	Cy5			
185724001	2002-10-25_Res_8-LA5Cy3_LACy5	mouse liver RNA A5	Cy3	○	○	◉
		mouse liver RNA A3	Cy5			
185724001	2002-10-9_Res1-Cy3LA3_Cy5LA5	mouse liver RNA A3	Cy3	○	○	◉
		mouse liver RNA A5	Cy5			
185724001	2002-10-9_Res2-Cy3LA3_Cy5LA5	mouse liver RNA A3	Cy3	○	○	◉
		mouse liver RNA A5	Cy5			

< Back R Reset > Continue

Figure 30 – Select Arrays for Comparison: Two Groups

Type II: Comparison of Samples in the Same Arrays

Choose this type of comparison and click on **Continue**, you will enter **Select Arrays for Comparison Analysis** page that allows you to select arrays for a channel-to-channel comparison (Figure 31). Note that you may make **single-sample comparison** by selecting a single array, or **multiple-sample comparison** by selecting multiple arrays. Because this comparison is between the two channels of arrays of interest, for Affymetrix chips, this type of comparison is not available. Click on **Reset** to cancel the current selections and re-select. Click on **Continue** to come to the next page where you are provided with a combination of comparison methods and parameters (Figure 32).



Select Arrays for Comparison Analysis

Please select arrays that you wish to compare, either *with* or *without* replicates. If there are replicates for both conditions, statistical significance will be evaluated using the methods on the next page.

Experiment	Array Name	Sample Name	Dye	Select	Not Select
185724001	2002-10-11_Res3-Cy3-Pool-vs-Cy5_LA5	mouse RNA pool A3	Cy3	○	◉
		mouse liver RNA A5	Cy5		
185724001	2002-10-11_Res4-Cy3-Pool-vs-Cy5_LA5	mouse RNA pool A3	Cy3	○	◉
		mouse liver RNA A5	Cy5		
185724001	2002-10-22_Res5-Cy3LA5Cy5pool	mouse liver RNA A5	Cy3	○	◉
		mouse RNA pool A3	Cy5		
185724001	2002-10-22_Res6-Cy3LA5Cy5pool	mouse liver RNA A5	Cy3	○	◉
		mouse RNA pool A3	Cy5		
185724001	2002-10-25_Res_7-LA5Cy3_LACy5	mouse liver RNA A5	Cy3	○	◉
		mouse liver RNA A3	Cy5		
185724001	2002-10-25_Res_8-LA5Cy3_LACy5	mouse liver RNA A5	Cy3	○	◉
		mouse liver RNA A3	Cy5		
185724001	2002-10-9_Res1-Cy3LA3_Cy5LA5	mouse liver RNA A3	Cy3	○	◉
		mouse liver RNA A5	Cy5		
185724001	2002-10-9_Res2-Cy3LA3_Cy5LA5	mouse liver RNA A3	Cy3	○	◉
		mouse liver RNA A5	Cy5		

< Back R Reset > Continue

Figure 31 – Select Arrays for comparison: Single Group

7.5.2. Select the Criteria for Differentially Expressed Genes

Type I: Comparison of Two Groups of Arrays

Once you have selected arrays that you wish to compare and clicked on **Continue** (Figure 30), you will enter the page where a combination of comparison methods/parameters are provided as criteria for differentially expressed gene(s) (Figure 32).

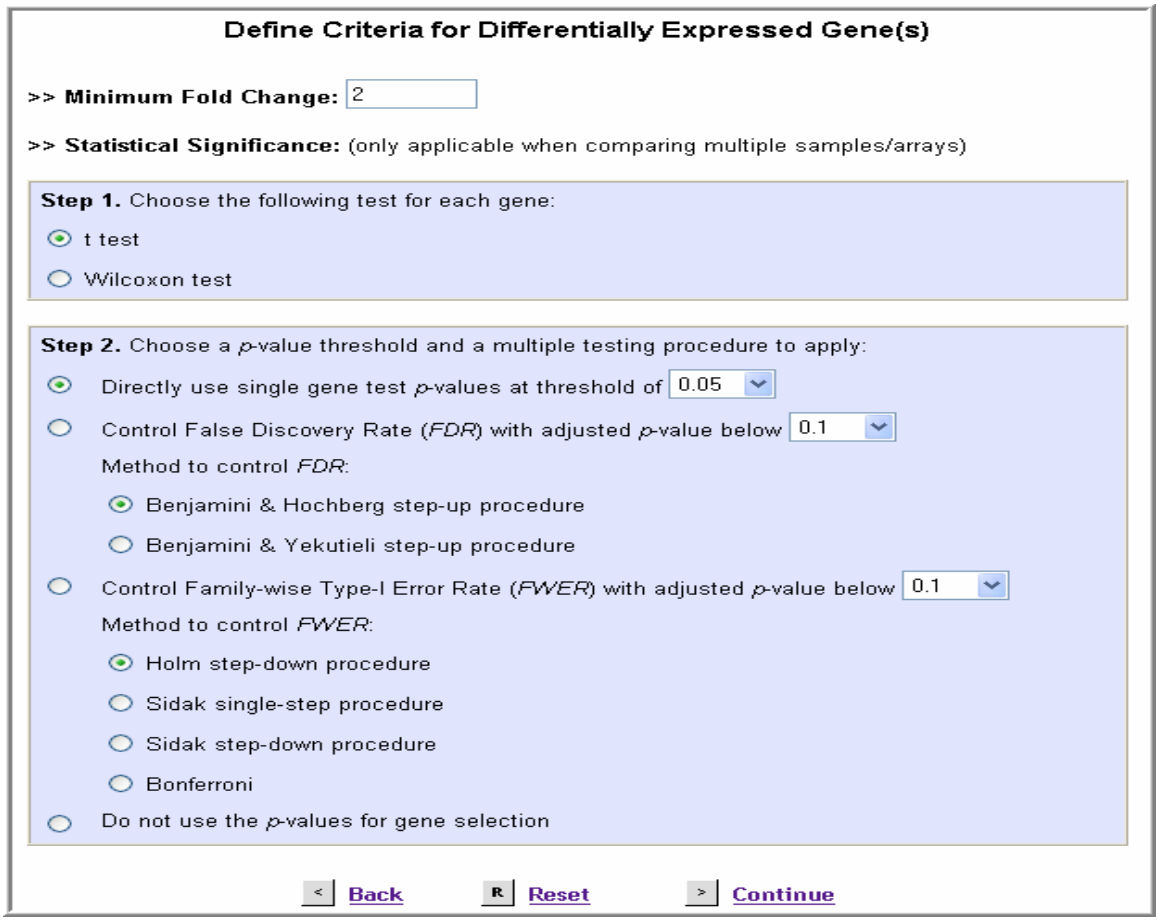


Figure 32 – Criteria for Differentially Expressed Genes

The user can choose either or both of the following criteria to select differentially expressed genes:

1. Specify a **Minimum Fold Change** between (group) A and B. Any genes that have fold changes greater between (group) A and B will be designated as differentially expressed by this criterion.
2. For **multiple-sample comparison**, the user can choose a type of test statistics, and a threshold for it, as selecting criteria. The **p-values** from

single test for each gene, or **adjusted p-values** computed with the selected multiple-testing method for control of false positive rate or family-wise error rate, can be used. The user can also choose not to use **p-value** for gene selection. For **single-sample comparison**, the statistical test procedures here are not applicable.

In addition, the following procedures are also available for **multiple-sample comparison**:

The magnitude of the differences between the geometric means of expression levels for each gene from the two groups will be computed. The significance of the differences is also evaluated, with two-sample **t test** with un-equal variance (Welch) or non-parametric **Wilcoxon** rank sum test (Mann-Whitney), as specified by the user.

Several methods to control the **false discovery rate (FDR)** in comparison of thousands of genes in microarray are provided. CEBS currently implements the **Benjamini & Hochberg** and **Benjamini & Yekutieli** procedures for strong control of the **false discovery rate (FDR)**. The **adjusted p-values** will be computed using the method selected by the user.

CEBS also provides means to control **family-wise Type-I error rate (FWER)**. The user can select **Holm**, **Sidak, or Bonferroni** procedure for strong control of the **family-wise Type I error rate**.

Type II: Comparison of Samples in the Same Arrays

Once you have selected arrays that you wish to compare and select **Continue** (Figure 31), you will now enter the page where a combination of comparison methods and parameters are provided as criteria for differentially expressed gene(s) (Figure 33).

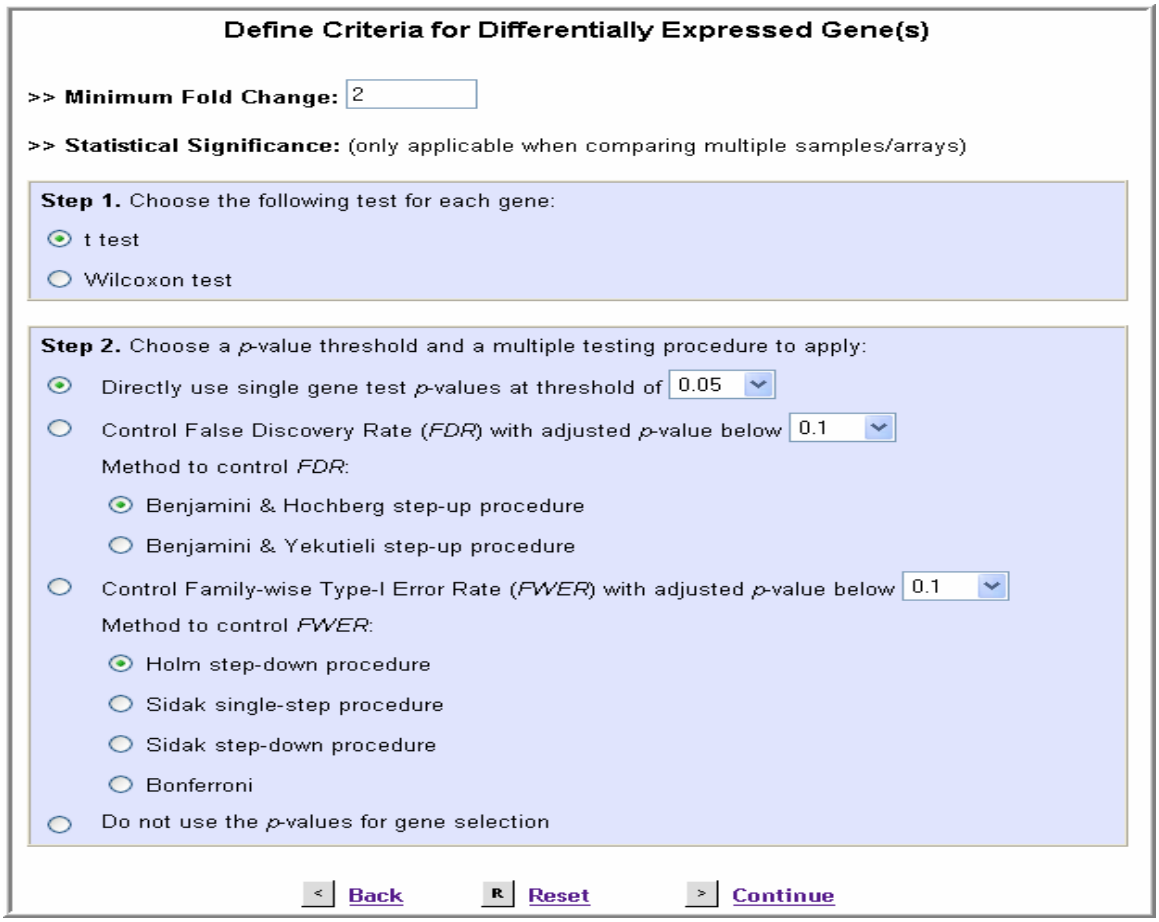


Figure 33 – Criteria for Differentially Expressed Genes

The user can choose either or both of the following criteria to select differentially expressed genes:

1. Specify a **Minimum Fold Change** between two channels. Any genes that have fold changes greater between two channels will be designated as differentially expressed by this criterion.
2. For **multiple-sample comparison**, the user can choose a type of test statistics, and a threshold for it, as selecting criteria. The **p-values** from single test for each gene, or **adjusted p-values** computed with the selected multiple-testing method for control of false positive rate or family-wise error rate, can be used. The user can also choose not to use **p-value** for gene selection. For **single-sample comparison**, the statistical test procedures here are not applicable.

In addition, the following procedures are also available for **multiple-sample comparison**:

The significance of the differences is evaluated, with one-sample **t test** or non-parametric one-sample **Wilcoxon** (signed rank) test, as specified by the user.

Several methods to control the **false discovery rate (FDR)** in comparison of thousands of genes in microarray are provided. CEBS currently implements the **Benjamini & Hochberg** and **Benjamini & Yekutieli** procedures for strong control of the **false discovery rate (FDR)**. The **adjusted p-values** will be computed using the method selected by the user.

CEBS also provides means to control **family-wise Type-I error rate (FWER)**. The user can select **Holm**, **Sidak**, or **Bonferroni** procedures for strong control of the **family-wise Type I error rate**.

7.5.3. References

For detailed information about the multiple testing procedures and their applications in microarray data analysis, see the following reference and documentation from the BioConductor documentation:

Y. Benjamini and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*. Vol. 57: 289-300.

Y. Benjamini and D. Yekutieli (2001). The control of the false discovery rate in multiple hypothesis testing under dependency. *Annals of Statistics*. Accepted.

S. Dudoit, J. P. Shaffer, and J. C. Boldrick (Submitted). Multiple hypothesis testing in microarray experiments.

Y. Ge, S. Dudoit, and T. P. Speed. Resampling-based multiple testing for microarray data hypothesis, Technical Report #633 of UCB Stat. http://www.stat.berkeley.edu/~gyc

Y. Hochberg (1988). A sharper Bonferroni procedure for multiple tests of significance, *Biometrika*. Vol. 75: 800-802.

S. Holm (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*. Vol. 6: 65-70.

7.6. Performing Biological Analysis of Gene Expression Data

Biological analyses are performed on the result of previous statistical analysis, e.g. the comparison analysis. Biological annotation and information about categories of biological activities (CBA) from different sources are incorporated during this stage of analysis to facilitate understanding of gene expression data.

CEBS provides the following options for a comprehensive biological analysis (Figure 34):

- Performing gene category analysis by BioCarta Pathways;
- Performing gene category analysis by KEGG Pathways;
- Performing gene category analysis by Gene Ontology (GO);
- Viewing expression report for all differentially expressed genes.

In these analyses, the gene level expression reports are presented together with gene annotations.

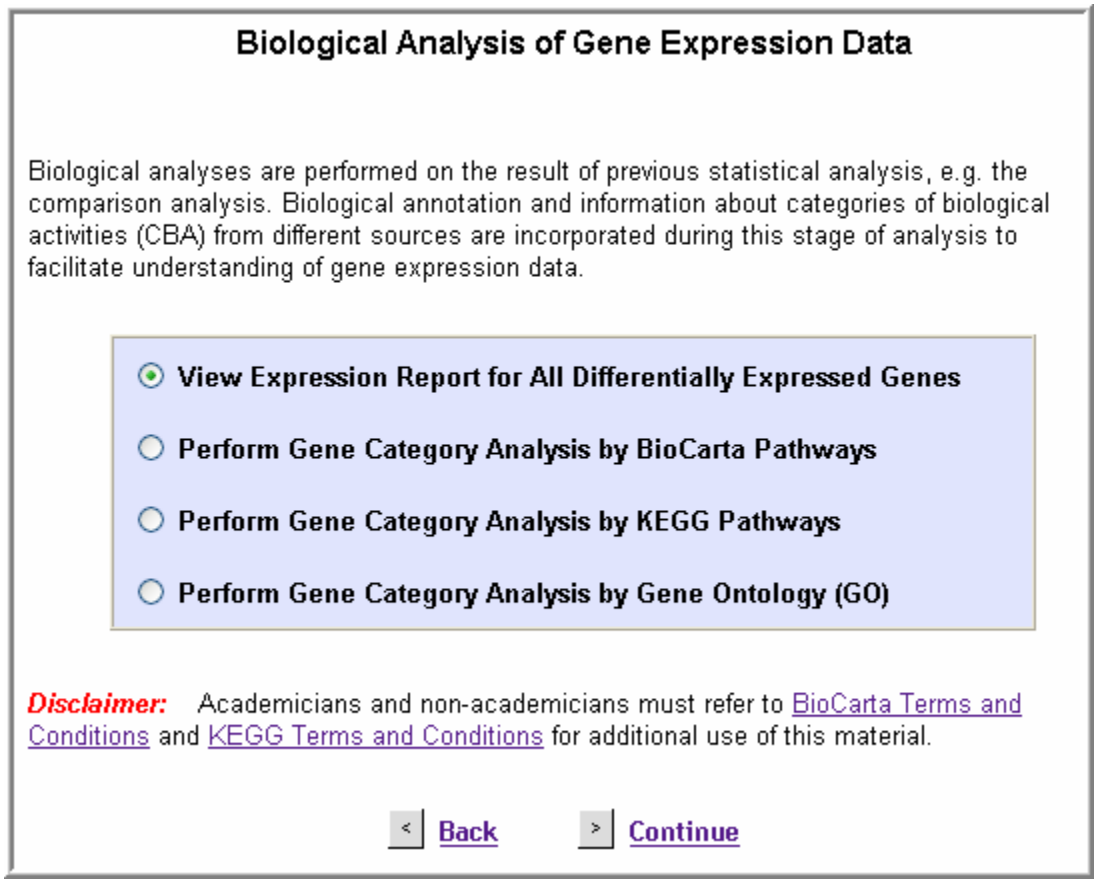
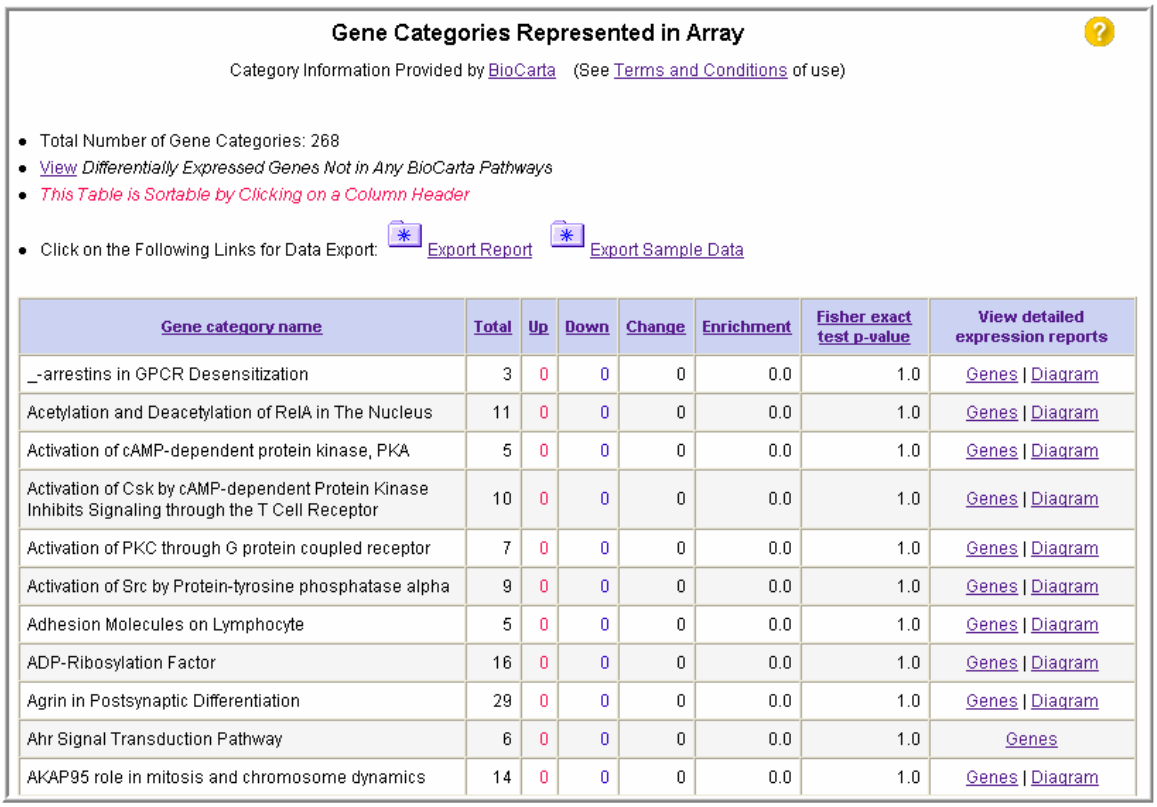


Figure 34 – Biological Analysis of Gene Expression Data

7.7. Perform Gene Category Analysis by BioCarta Pathways:

7.7.1. Gene Categories Represented in Array

Choose **Perform Gene Category Analysis by BioCarta Pathways** to view the **Gene Categories Represented in Array** page that displays a list of BioCarta pathways and statistical information for gene expression on the pathway level (Figure 35).



Gene Categories Represented in Array

Category Information Provided by BioCarta (See Terms and Conditions of use)

- Total Number of Gene Categories: 268
- View *Differentially Expressed Genes Not in Any BioCarta Pathways*
- *This Table is Sortable by Clicking on a Column Header*
- Click on the Following Links for Data Export: Export Report Export Sample Data

Gene category name	Total	Up	Down	Change	Enrichment	Fisher exact test p-value	View detailed expression reports
_-arrestins in GPCR Desensitization	3	0	0	0	0.0	1.0	Genes | Diagram
Acetylation and Deacetylation of RelA in The Nucleus	11	0	0	0	0.0	1.0	Genes | Diagram
Activation of cAMP-dependent protein kinase, PKA	5	0	0	0	0.0	1.0	Genes | Diagram
Activation of Csk by cAMP-dependent Protein Kinase Inhibits Signaling through the T Cell Receptor	10	0	0	0	0.0	1.0	Genes | Diagram
Activation of PKC through G protein coupled receptor	7	0	0	0	0.0	1.0	Genes | Diagram
Activation of Src by Protein-tyrosine phosphatase alpha	9	0	0	0	0.0	1.0	Genes | Diagram
Adhesion Molecules on Lymphocyte	5	0	0	0	0.0	1.0	Genes | Diagram
ADP-Ribosylation Factor	16	0	0	0	0.0	1.0	Genes | Diagram
Agrin in Postsynaptic Differentiation	29	0	0	0	0.0	1.0	Genes | Diagram
Ahr Signal Transduction Pathway	6	0	0	0	0.0	1.0	Genes
AKAP95 role in mitosis and chromosome dynamics	14	0	0	0	0.0	1.0	Genes | Diagram

Figure 35 – BioCarta Pathways Page

Gene Category Name column displays the pathway name. **Total** column displays the number of all the design elements for which the represented genes are in the specific pathway. **Up**, **Down** and **Change** columns represent the number of up-regulated, down-regulated and changed gene expressions for the comparison criteria you used. The relative proportion of changed genes in a gene category as compared to overall proportion of changed genes among the total number of genes with available expression value (excluding genes with missing expression data) is presented as an **Enrichment** factor. If the enrichment factor is greater than 1, the gene category has an enrichment of changed compared to all genes, while an enrichment factor less than 1 means

the gene category has a depletion in the changed gene. A two-sided **Fisher Exact Test p-value** is also provided as indication of the probability of observing such a number of changed genes in the pathway by chance (Figure 35).

Click on **Gene Category Name** column header to sort the pathways by pathway name alphabetically. Click on **Total**, **Up**, **Down**, **Change**, **Enrichment** column headers to sort the pathways by the specific column name in a descending order. Click on **Fisher Exact Test p-value** to sort the pathways in an ascending order.

For example, when you click the **Change** column header, the pathways are displayed in the descending order of the number of changed expressions. At this time, the pathways with differentially expressed genes, i.e. the pathways with a non-zero number in the **Change** column, will be distinguished from those without any changed gene expressions.

Select the link on the page to view the genes that have shown differential expression for the criteria you specified and are not in any BioCarta pathways. See **Differentially Expressed Genes Not in BioCarta Pathways** for details, Section 7.7.4.

7.7.2. Pathway Diagram

Pathway Diagram Page provides a re-engineered graphical representation of genes in a particular category of biological activity based on BioCarta pathways (Figure 36). It also incorporates into the pathway diagram the information about the gene expression, particularly the results from comparison between two biological states as chosen by the user.

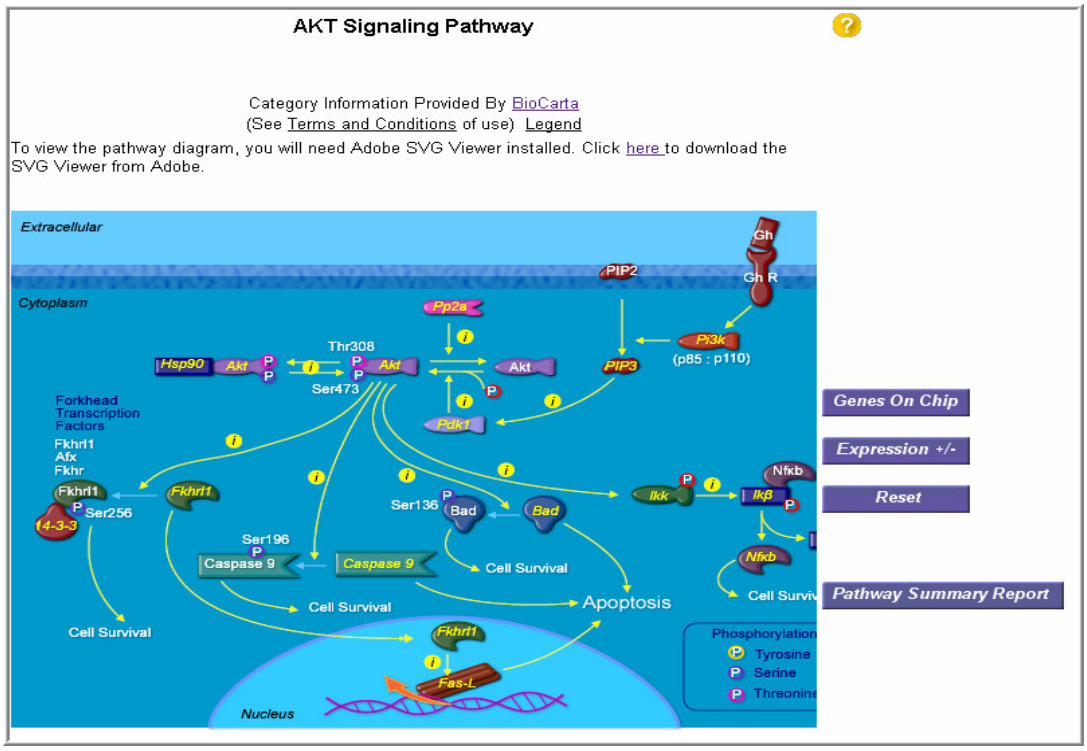


Figure 36 – BioCarta Pathway Diagram page

Click on a gene displayed on the pathway diagram to invoke the **Gene Info Page** (Figure 55) for more detailed information of the selected gene.

Select the ***Genes on Chip*** button to view the genes that are represented in the chip for the pathway selected highlighted in pink (Figure 37).

Select the ***Expression +/-*** button to view the differentially expressed genes highlighted in distinct colors: red for up-regulated genes, blue for down-regulated genes and green for those genes represented by multiple features (Figure 38).

Select the ***Reset*** to reset the pathway diagram (Figure 36).

Select the ***Pathway Summary Report*** to view the summary data of gene expression for the pathway selected (Figure 39).

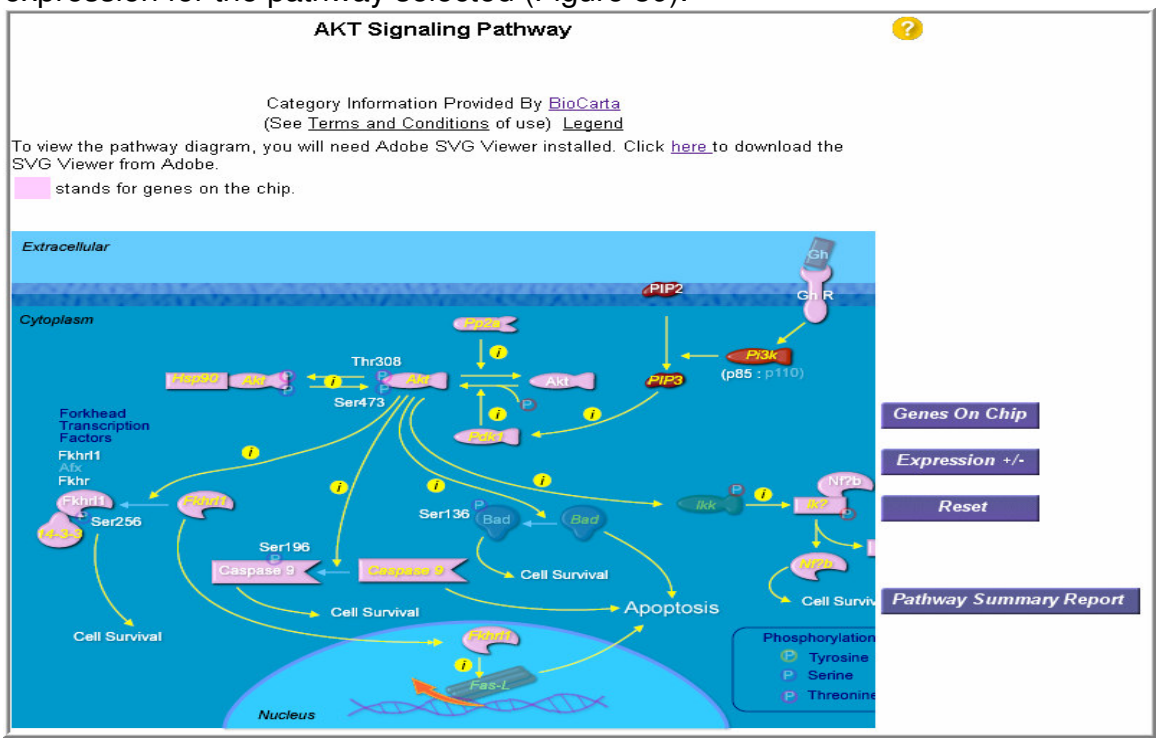


Figure 37 – BioCarta: Genes on Chip

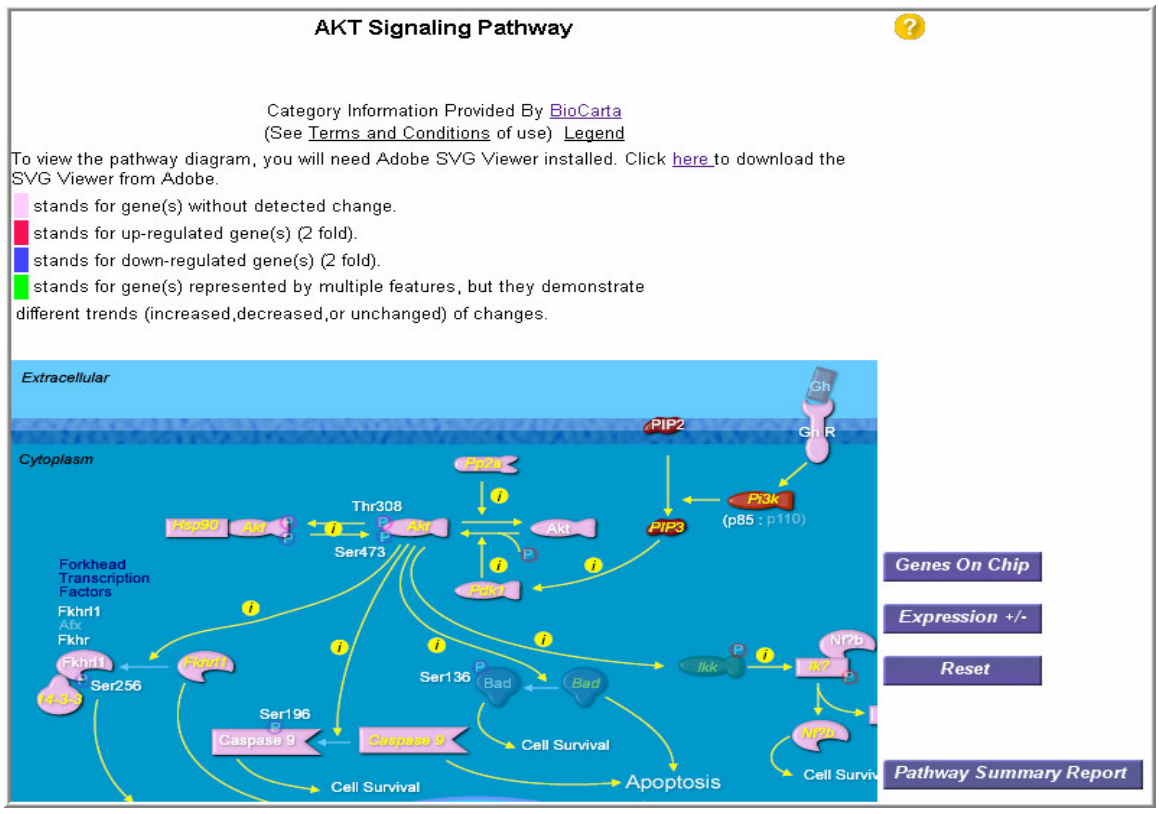


Figure 38 – BioCarta Pathway Summary Report

7.7.3. Expression Report for Genes in the Pathway

Expression Report for Genes in the Pathway displays a summary of the expression of genes in the selected pathway, including up-regulated, down-regulated and non-differentially expressed genes (Figure 39).

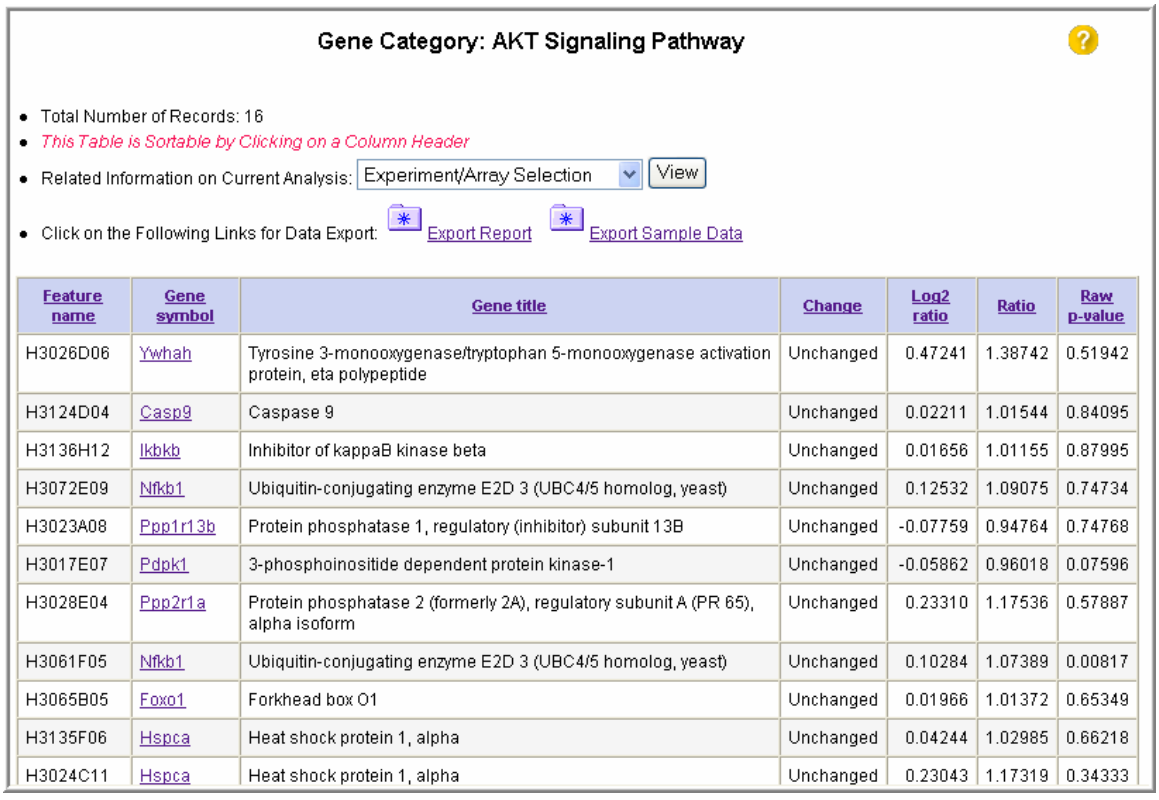
For two-channel arrays, the **Feature Name** represents the name of reporter. The expression levels are represented by **log2 ratios** (sample-sample comparisons), or **mean log2 ratios** (group-group comparisons). For Affymetrix arrays, the **Probe Set** represents the name of reporter. **log2 intensities** are used for sample-sample comparisons, while **mean log2 intensities** are used for group-group comparisons. The **Gene Symbol** and **Gene Title** represent the name and title of the target gene, respectively. The **Change** column is used to indicate whether the gene expression is up- or down-regulated, by the criteria selected by the user. **log2(A)-log2(B)** represents the difference between (group) A and B, while **Log2 ratio** is used for the difference between two channels. **A/B** represents the fold change between (group) A and B, while **Ratio** is used for the fold change between two channels. A **Raw p-value** is displayed to indicate the

confidence level of differential expression between the two groups of arrays, based on the statistical test for the data on the particular gene. An **Adjusted p-value** is derived from the single test p-value, through applying a selected multiple testing procedure to control the false discovery rate or family-wise error rate.

Click on a **Gene Symbol** on the list to (Figure 39) view the detailed gene information on **Gene Info Page** (Figure 55).

Click on a column header to sort the pathways in order of that column. For example, select the **Raw p-value** column to display the pathways in an ascending order based on the value of the raw p-value.

In addition to the gene expression data, the report also provides the related information on current analysis: **Experiment/Array Selection** (Figure 40) and **Options Used for Comparison** (Figure 41). Choose to review the information about the above items by clicking on **View** button.



Gene Category: AKT Signaling Pathway

- Total Number of Records: 16
- *This Table is Sortable by Clicking on a Column Header*
- Related Information on Current Analysis: Experiment/Array Selection View
- Click on the Following Links for Data Export: Export Report Export Sample Data

Feature name	Gene symbol	Gene title	Change	Log2 ratio	Ratio	Raw p-value
H3026D06	Ywhah	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	Unchanged	0.47241	1.38742	0.51942
H3124D04	Casp9	Caspase 9	Unchanged	0.02211	1.01544	0.84095
H3136H12	Ikbkb	Inhibitor of kappaB kinase beta	Unchanged	0.01656	1.01155	0.87995
H3072E09	Nfkb1	Ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	Unchanged	0.12532	1.09075	0.74734
H3023A08	Ppp1r13b	Protein phosphatase 1, regulatory (inhibitor) subunit 13B	Unchanged	-0.07759	0.94764	0.74768
H3017E07	Pdpk1	3-phosphoinositide dependent protein kinase-1	Unchanged	-0.05862	0.96018	0.07596
H3028E04	Ppp2r1a	Protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), alpha isoform	Unchanged	0.23310	1.17536	0.57887
H3061F05	Nfkb1	Ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	Unchanged	0.10284	1.07389	0.00817
H3065B05	Foxo1	Forkhead box O1	Unchanged	0.01966	1.01372	0.65349
H3135F06	Hspca	Heat shock protein 1, alpha	Unchanged	0.04244	1.02985	0.66218
H3024C11	Hspca	Heat shock protein 1, alpha	Unchanged	0.23043	1.17319	0.34333

Figure 39 – Expression Report for Genes

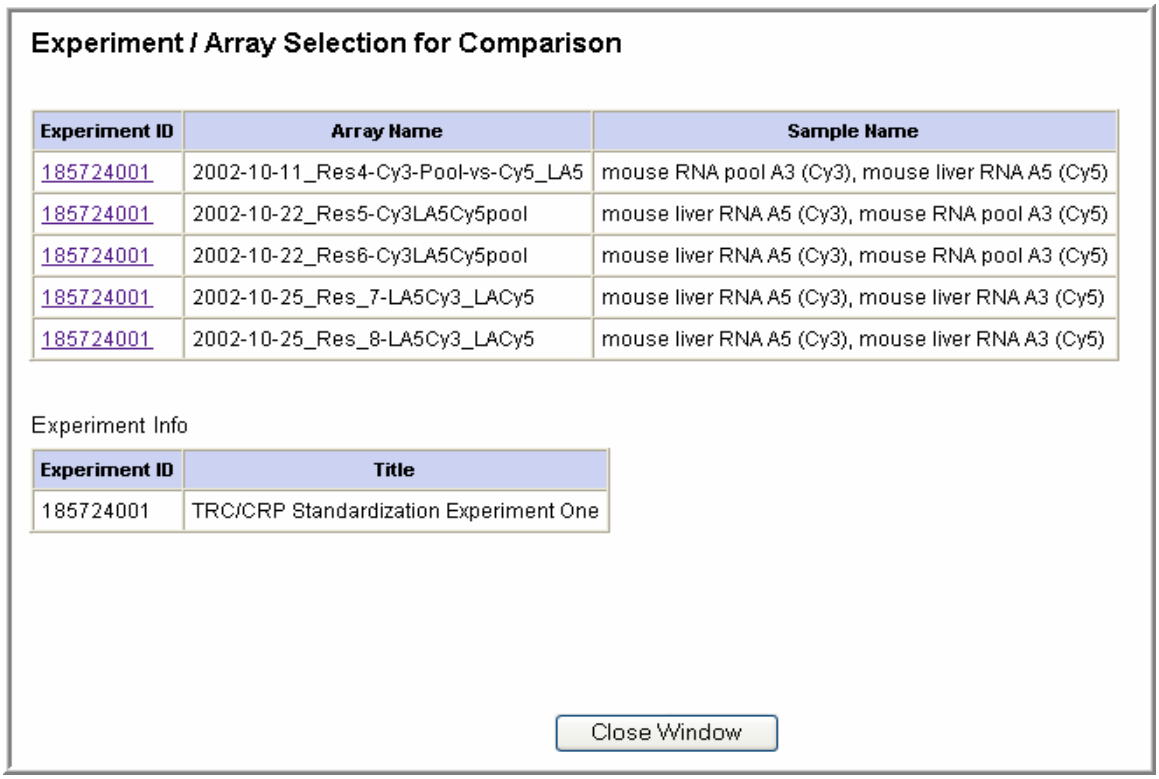


Figure 40 – The Experiment & Array Currently Selected for Comparison

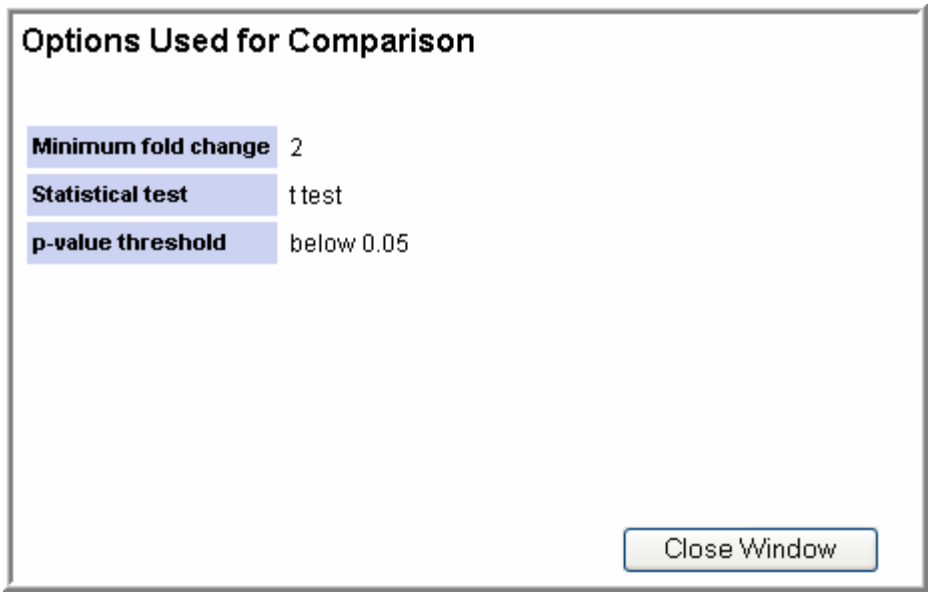


Figure 41 – The Options Currently used for Comparison

7.7.4. Differentially Expressed Genes Not in BioCarta Pathways

Differentially Expressed Genes Not in BioCarta Pathways report displays a list of genes that are differentially expressed based on the comparison criteria you selected which are not in any BioCarta pathways (Figure 42).

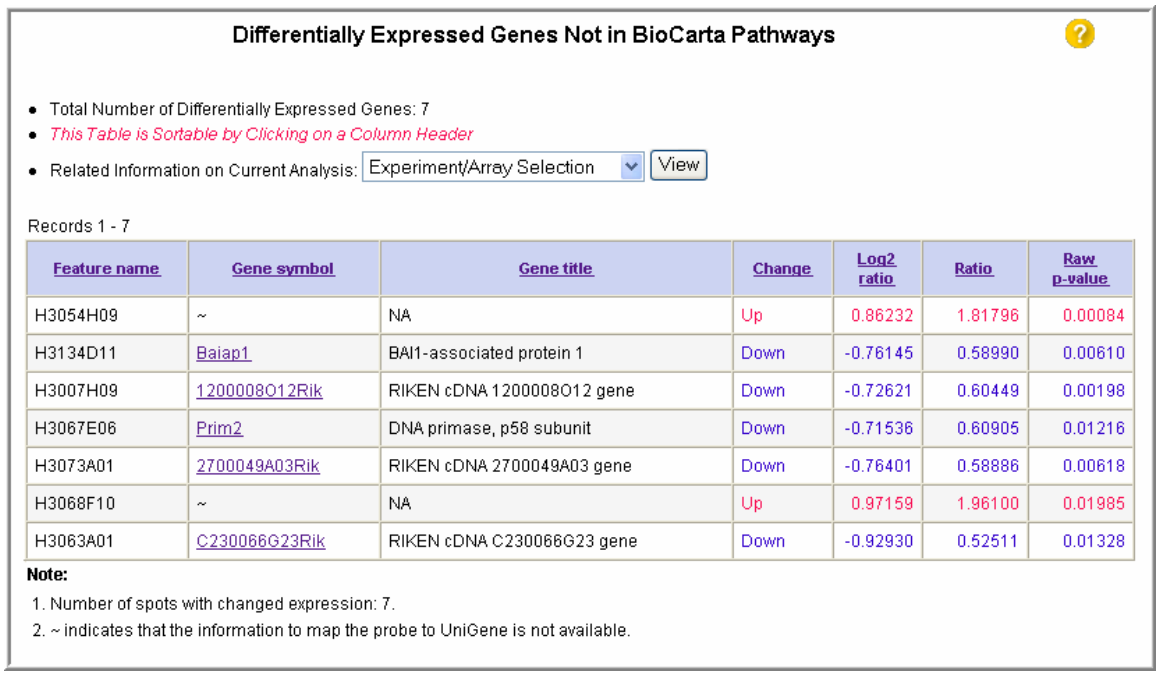
For two-channel arrays, the **Feature Name** represents the name of reporter. The expression levels are represented by **log2 ratios** (sample-sample comparisons), or **mean log2 ratios** (group-group comparisons). For Affymetrix arrays, the **Probe Set** represents the name of reporter. **log2 intensities** are used for sample-sample comparisons, while **mean log2 intensities** are used for group-group comparisons. The **Gene Symbol** and **Gene Title** represent the name and title of target gene, respectively. The **Change** column is used to indicate whether the gene expression is up- or down-regulated using the criteria selected by the user. **log2(A)-log2(B)** represents the difference between (group) A and B, while **Log2 ratio** is used for the difference between two channels. **A/B** represents the fold change between (group) A and B, while **Ratio** is used for the fold change between two channels. A **Raw p-value** is displayed to indicate the confidence level of differential expression between the two groups of arrays, based on the statistical test for the data on the particular gene. An **Adjusted p-value** is derived from the single test p-value, through applying a selected multiple testing procedure to control the false discovery rate or family-wise error rate.

Select the the name of a **Gene Symbol** (Figure 42) on the list to view the detailed gene information on **Gene Info Page** (Figure 55).

Select the column header to sort the pathways in order of that column. For example, select the **Raw p-value** column to display the pathways in an ascending order based on the value of the raw p-value.

The report has a function for pagination. If the number of genes is large, the report is split into smaller pages with 200 genes per page. Users can navigate through the pages by choosing a page number to go to.

In addition to the gene expression data, the report also provides the related information on current analysis: **Experiment/Array Selection** (Figure 40) and **Options Used for Comparison** (Figure 41). Choose to review the information about the above items by clicking on **View** button.



Differentially Expressed Genes Not in BioCarta Pathways

- Total Number of Differentially Expressed Genes: 7
- *This Table is Sortable by Clicking on a Column Header*
- Related Information on Current Analysis: Experiment/Array Selection View

Records 1 - 7

Feature name	Gene symbol	Gene title	Change	Log2 ratio	Ratio	Raw p-value
H3054H09	~	NA	Up	0.86232	1.81796	0.00084
H3134D11	Baiap1	BAI1-associated protein 1	Down	-0.76145	0.58990	0.00610
H3007H09	1200008O12Rik	RIKEN cDNA 1200008O12 gene	Down	-0.72621	0.60449	0.00198
H3067E06	Prim2	DNA primase, p58 subunit	Down	-0.71536	0.60905	0.01216
H3073A01	2700049A03Rik	RIKEN cDNA 2700049A03 gene	Down	-0.76401	0.58886	0.00618
H3068F10	~	NA	Up	0.97159	1.96100	0.01985
H3063A01	C230066G23Rik	RIKEN cDNA C230066G23 gene	Down	-0.92930	0.52511	0.01328

Note:

1. Number of spots with changed expression: 7.
2. ~ indicates that the information to map the probe to UniGene is not available.

Figure 42 – Differentially Expressed genes not in BioCarta Pathway

7.8. Perform Gene Category Analysis by KEGG Pathways:

When you choose **Perform Gene Category Analysis by KEGG Pathways** (Figure 34), you will enter the main page for KEGG pathway analysis (Figure 43).

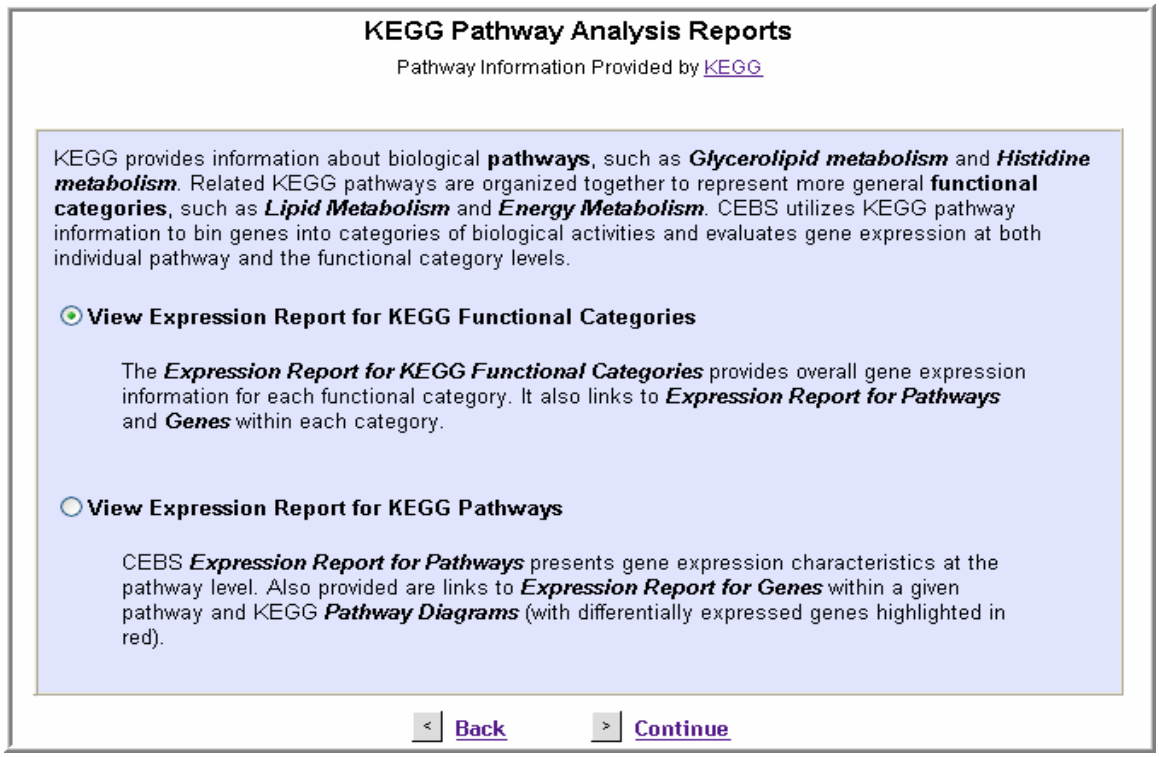


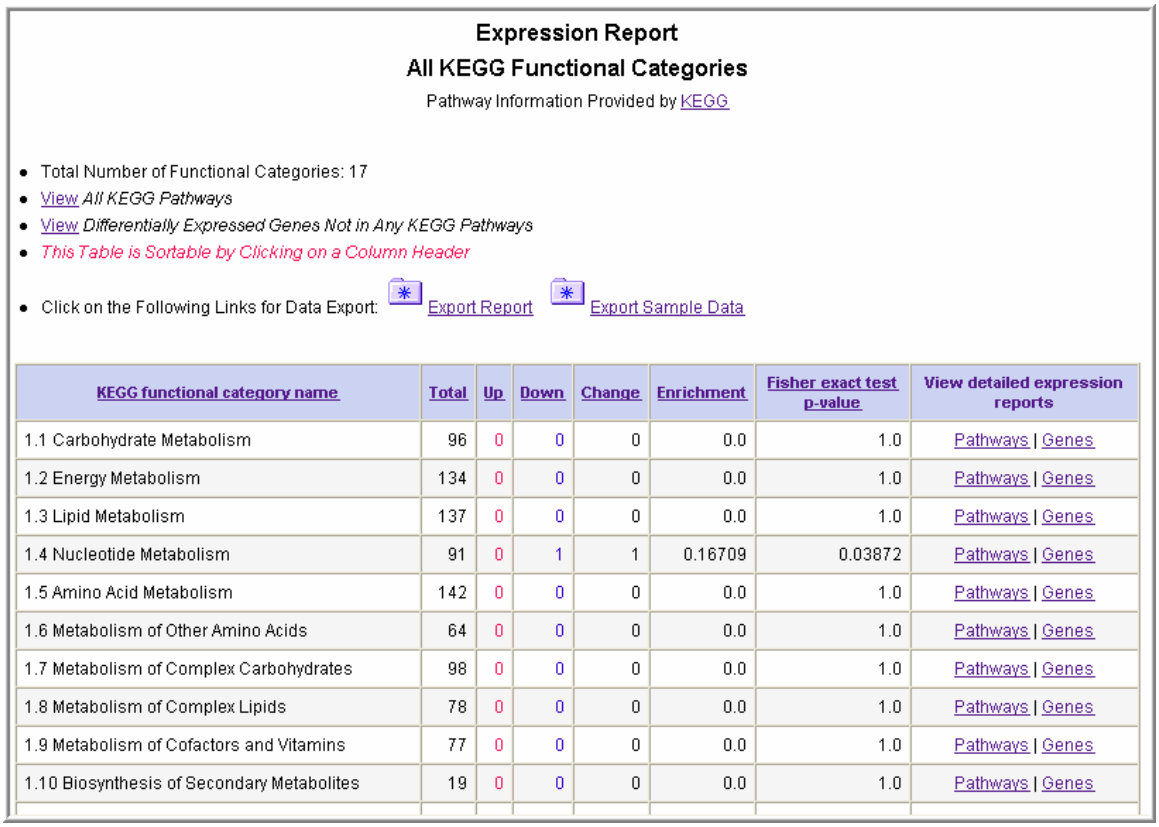
Figure 43 – KEGG Pathway Analysis Reports

KEGG provides information about biological **pathways**, such as ***Glycerolipid metabolism*** and ***Histidine metabolism***. Related KEGG pathways are organized together to represent more general **functional categories**, such as ***Lipid Metabolism*** and ***Energy Metabolism***. CEBS utilizes KEGG pathway information to bin genes into categories of biological activities and evaluates gene expression at both individual pathway and the functional category levels. You can choose to view (1) **Expression Report for KEGG Functional Categories**, or (2) **Expression Report for KEGG Pathways**.

Expression Report for KEGG Functional Categories: This report (Figure 44) provides overall gene expression information for each functional category. It also links to **Expression Report for KEGG Pathways** and **Genes** within each category.

KEGG Functional Category Name column displays the functional category name. **Total** column displays the number of all the design elements for which the represented genes are in the specific pathway. **Up**, **Down** and **Change** columns represent the number of up-regulated, down-regulated and changed gene expressions for the comparison criteria you used. The relative proportion of changed genes in a gene category as compared to overall proportion of changed genes among the total number of genes with available expression value (excluding genes with missing expression data) is presented as an **Enrichment**

factor. If the enrichment factor is greater than 1, the gene category has an enrichment of changed compared to all genes, while an enrichment factor less than 1 means the gene category has a depletion in the changed gene. A two-sided **Fisher Exact Test p-value** is also provided as indication of the probability of observing such a number of changed genes in the KEGG functional categories by chance.



Expression Report

All KEGG Functional Categories

Pathway Information Provided by KEGG

- Total Number of Functional Categories: 17
- View *All KEGG Pathways*
- View *Differentially Expressed Genes Not in Any KEGG Pathways*
- *This Table is Sortable by Clicking on a Column Header*
- Click on the Following Links for Data Export: Export Report Export Sample Data

KEGG functional category name	Total	Up	Down	Change	Enrichment	Fisher exact test p-value	View detailed expression reports
1.1 Carbohydrate Metabolism	96	0	0	0	0.0	1.0	Pathways | Genes
1.2 Energy Metabolism	134	0	0	0	0.0	1.0	Pathways | Genes
1.3 Lipid Metabolism	137	0	0	0	0.0	1.0	Pathways | Genes
1.4 Nucleotide Metabolism	91	0	1	1	0.16709	0.03872	Pathways | Genes
1.5 Amino Acid Metabolism	142	0	0	0	0.0	1.0	Pathways | Genes
1.6 Metabolism of Other Amino Acids	64	0	0	0	0.0	1.0	Pathways | Genes
1.7 Metabolism of Complex Carbohydrates	98	0	0	0	0.0	1.0	Pathways | Genes
1.8 Metabolism of Complex Lipids	78	0	0	0	0.0	1.0	Pathways | Genes
1.9 Metabolism of Cofactors and Vitamins	77	0	0	0	0.0	1.0	Pathways | Genes
1.10 Biosynthesis of Secondary Metabolites	19	0	0	0	0.0	1.0	Pathways | Genes

Figure 44 – Expression Report for KEGG Functional Categories

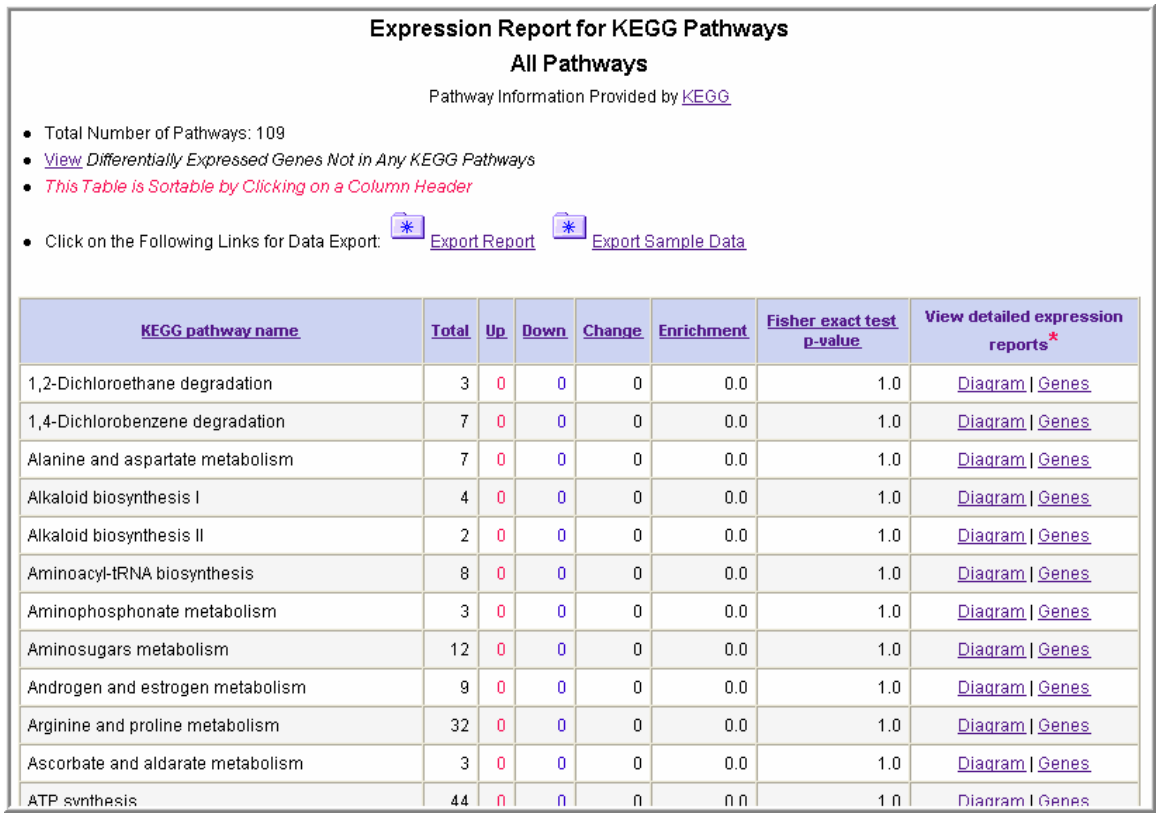
Click on **KEGG Functional Category Name** column header to sort the functional categories by name alphabetically. Click on **Total**, **Up**, **Down**, **Change**, **Enrichment** to sort the pathways by the specific column name in descending order. Click on **Fisher Exact Test p-value** to sort the pathways in ascending order.

For example, when you click on **Enrichment** column header, the pathways are displayed in the descending order of the number of changed expressions. At this time, the pathways with differentially expressed genes, i.e. the pathways with a non-zero number in the **Enrichment** column, will be distinguished from those without any changed gene expressions.

From the KEGG functional category level, there are two ways to view expression details: view a table of KEGG pathways within a given functional category (by

clicking on **Pathways**) (Figure 46), or view a table of genes with expression information (**KEGG Expression Report for Genes** by clicking on **Genes**) (Figure 48).

Expression Report for KEGG Pathways: CEBS **Expression Report for KEGG Pathways** presents gene expression characteristics at the pathway level. Also provided are links to ***Expression Report for Genes*** within a given pathway and **KEGG Pathway Diagrams** (with differentially expressed genes highlighted in red). You can choose to view **Expression Report for KEGG Pathways** for all available KEGG pathways (Figure 46), or for KEGG pathways within a functional category (Figure 46).



Expression Report for KEGG Pathways

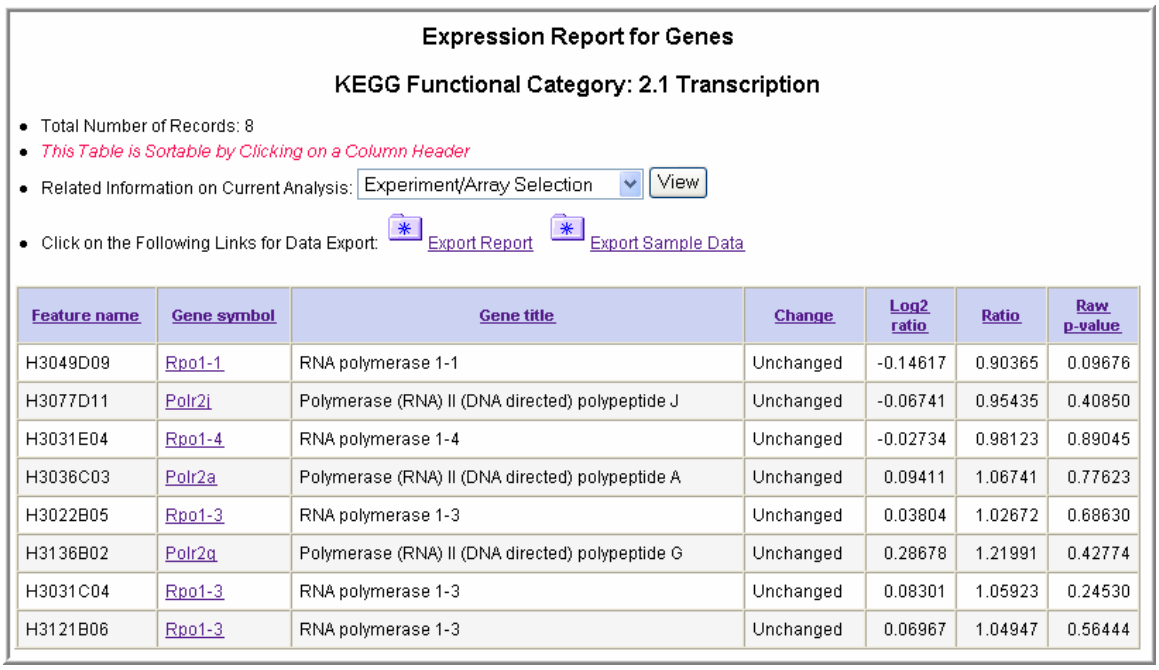
All Pathways

Pathway Information Provided by KEGG

- Total Number of Pathways: 109
- View *Differentially Expressed Genes Not in Any KEGG Pathways*
- *This Table is Sortable by Clicking on a Column Header*
- Click on the Following Links for Data Export: Export Report Export Sample Data

KEGG pathway name	Total	Up	Down	Change	Enrichment	Fisher exact test p-value	View detailed expression reports*
1,2-Dichloroethane degradation	3	0	0	0	0.0	1.0	Diagram | Genes
1,4-Dichlorobenzene degradation	7	0	0	0	0.0	1.0	Diagram | Genes
Alanine and aspartate metabolism	7	0	0	0	0.0	1.0	Diagram | Genes
Alkaloid biosynthesis I	4	0	0	0	0.0	1.0	Diagram | Genes
Alkaloid biosynthesis II	2	0	0	0	0.0	1.0	Diagram | Genes
Aminoacyl-tRNA biosynthesis	8	0	0	0	0.0	1.0	Diagram | Genes
Aminophosphonate metabolism	3	0	0	0	0.0	1.0	Diagram | Genes
Aminosugars metabolism	12	0	0	0	0.0	1.0	Diagram | Genes
Androgen and estrogen metabolism	9	0	0	0	0.0	1.0	Diagram | Genes
Arginine and proline metabolism	32	0	0	0	0.0	1.0	Diagram | Genes
Ascorbate and aldarate metabolism	3	0	0	0	0.0	1.0	Diagram | Genes
ATP synthesis	44	0	0	0	0.0	1.0	Diagram | Genes

Figure 45 – Expression Report for KEGG Pathway for all available KEGG Pathways



Expression Report for Genes

KEGG Functional Category: 2.1 Transcription

- Total Number of Records: 8
- *This Table is Sortable by Clicking on a Column Header*
- Related Information on Current Analysis: Experiment/Array Selection View
- Click on the Following Links for Data Export: Export Report Export Sample Data

Feature name	Gene symbol	Gene title	Change	Log2 ratio	Ratio	Raw p-value
H3049D09	Rpo1-1	RNA polymerase 1-1	Unchanged	-0.14617	0.90365	0.09676
H3077D11	Polr2j	Polymerase (RNA) II (DNA directed) polypeptide J	Unchanged	-0.06741	0.95435	0.40850
H3031E04	Rpo1-4	RNA polymerase 1-4	Unchanged	-0.02734	0.98123	0.89045
H3036C03	Polr2a	Polymerase (RNA) II (DNA directed) polypeptide A	Unchanged	0.09411	1.06741	0.77623
H3022B05	Rpo1-3	RNA polymerase 1-3	Unchanged	0.03804	1.02672	0.68630
H3136B02	Polr2g	Polymerase (RNA) II (DNA directed) polypeptide G	Unchanged	0.28678	1.21991	0.42774
H3031C04	Rpo1-3	RNA polymerase 1-3	Unchanged	0.08301	1.05923	0.24530
H3121B06	Rpo1-3	RNA polymerase 1-3	Unchanged	0.06967	1.04947	0.56444

Figure 46 – Expression Report for KEGG Pathways with in a Functional Category

KEGG Pathway Name column displays the pathway name. **Total** column displays the number of all the design elements for which the represented genes are in the specific pathway. **Up**, **Down** and **Change** columns represent the number of up-regulated, down-regulated and changed gene expressions for the comparison criteria you used. The relative proportion of changed genes in a gene category as compared to overall proportion of changed genes among the total number of genes with available expression value (excluding genes with missing expression data) is presented as an **Enrichment** factor. If the enrichment factor is greater than 1, the gene category has an enrichment of changed compared to all genes, while an enrichment factor less than 1 means the gene category has a depletion in the changed gene. A two-sided **Fisher Exact Test p-value** is also provided as indication of the probability of observing such a number of changed genes in the KEGG pathway by chance.

From the KEGG pathway level expression report, there are two ways to view details for a KEGG pathway: graphically through the **KEGG Pathway Diagram** (by clicking on **Diagram**), or view **KEGG Expression Report for Genes**, a table of genes with expression information (by clicking on **Genes**).

KEGG Pathway Diagram (external): (Figure 47)

- The KEGG pathway diagrams are provided via external link to the KEGG web site.
- The differentially expressed gene(s) are posted to the KEGG site to be highlighted (in red) in the diagram. However, not every gene can be

manipulated in color in KEGG diagrams, thus some of the differentially expressed genes may not be highlighted.

- For expression characteristics of all genes in a pathway, please refer to the gene report for that pathway.
- Please note that some of the pathway diagrams for a given species may not be available from KEGG. In that case, you will get a "can't find pathway map" notice from KEGG.

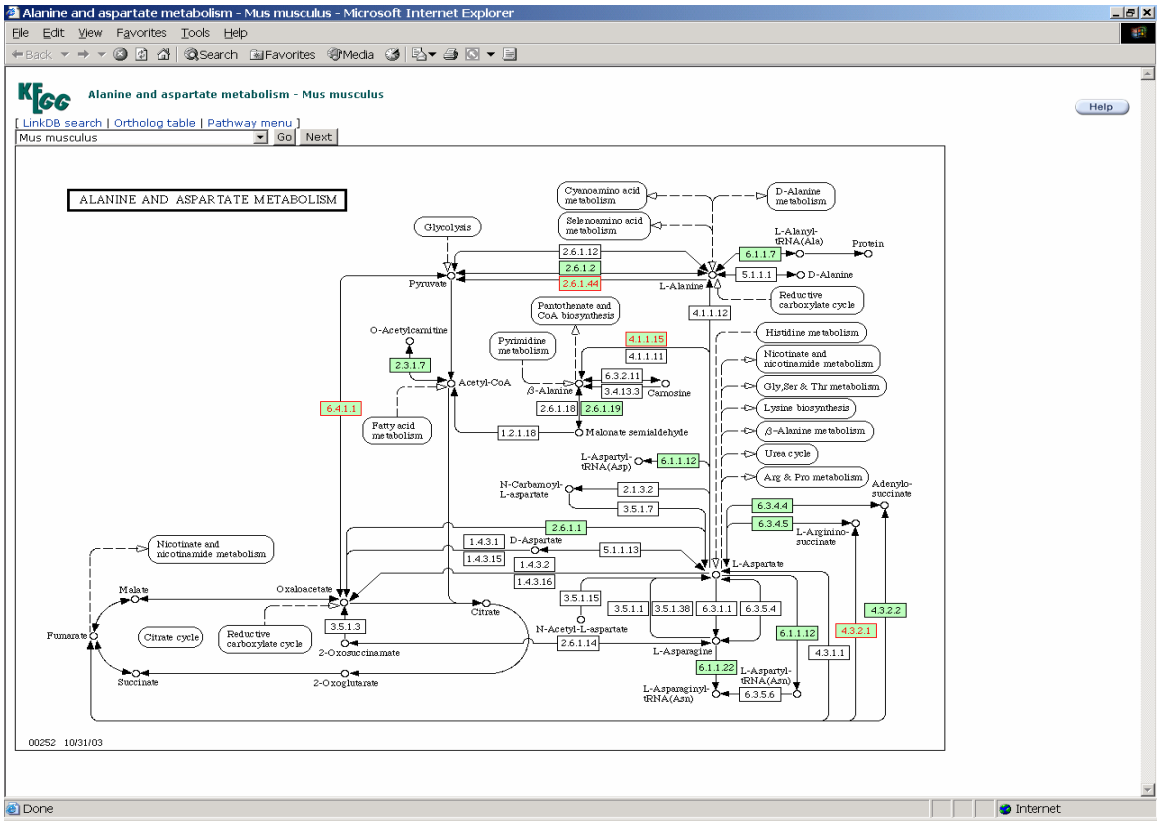
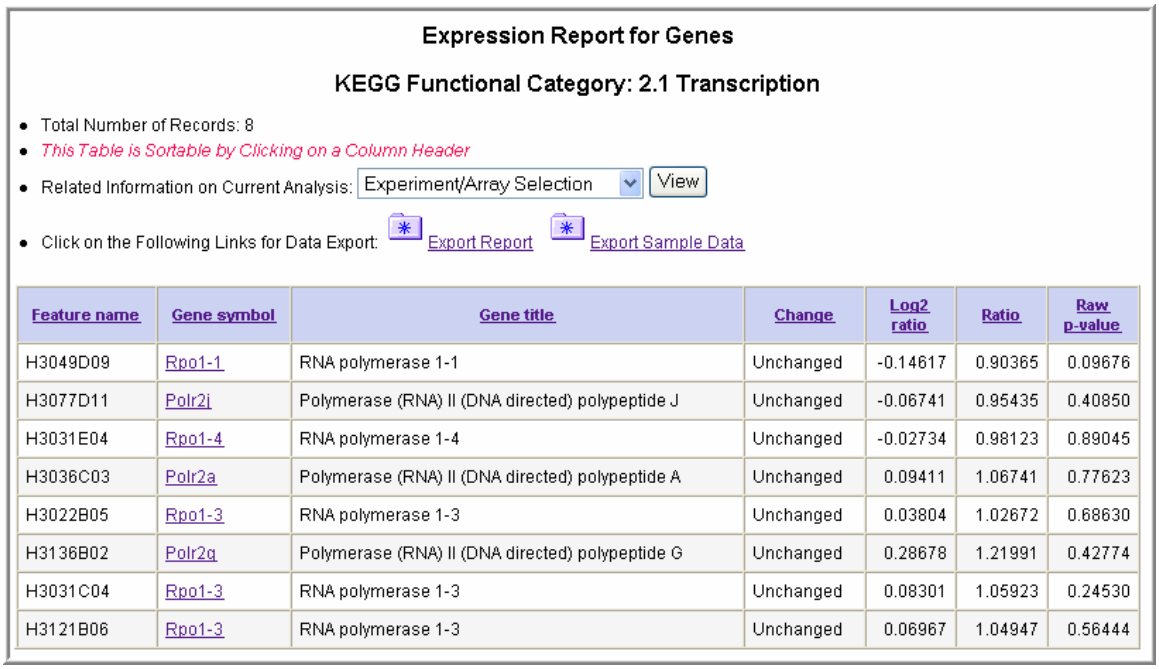


Figure 47 – KEGG Pathway Diagram

KEGG Expression Report for Genes: This is expression report at gene level for a KEGG pathway (Figure 49), or a more general, functional category (Figure 46). This report is similar to **Gene Expression Report for BioCarta Pathways** (for detailed description, please refer to the corresponding section of the user guide for BioCarta pathways). Also available is a list of genes and expression information for genes that are not in any KEGG pathways (Figure 50).

Click on a column header to sort the pathways in order of that column. For example, click on the **Raw p-value** column to display the pathways in an ascending order based on the value of the raw p-value.

In addition to the gene expression data, the report also provides the related information on current analysis: **Experiment/Array Selection** (Figure 40) and **Options Used for Comparison** (Figure 41). Choose to review the information about the above items by clicking on **View** button.



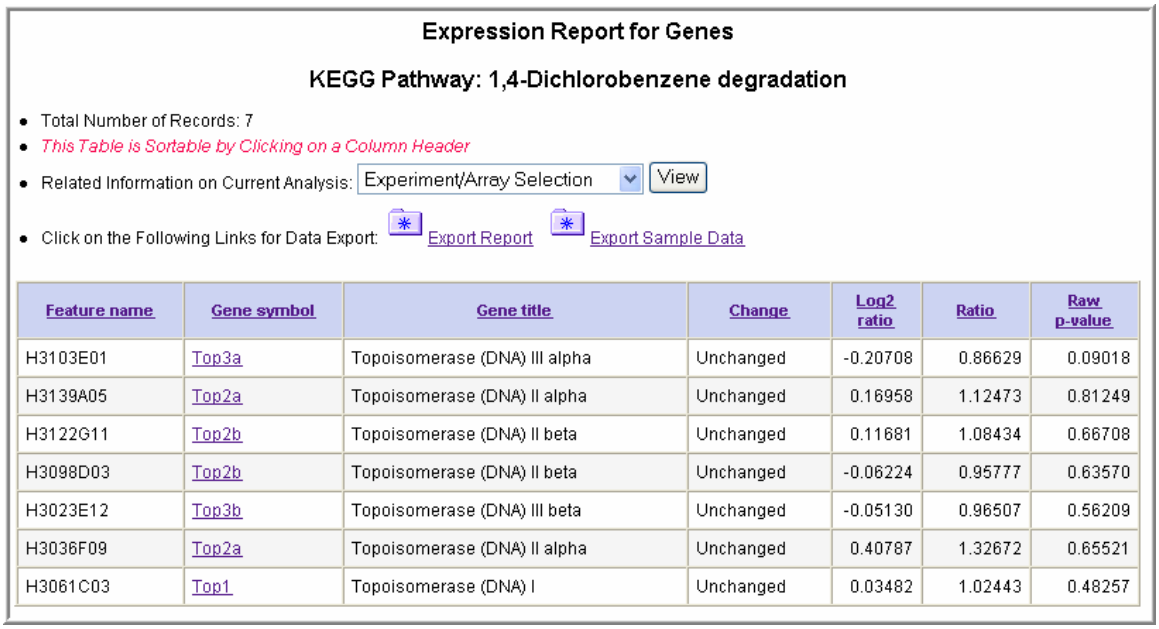
Expression Report for Genes

KEGG Functional Category: 2.1 Transcription

- Total Number of Records: 8
- *This Table is Sortable by Clicking on a Column Header*
- Related Information on Current Analysis: Experiment/Array Selection View
- Click on the Following Links for Data Export: Export Report Export Sample Data

Feature name	Gene symbol	Gene title	Change	Log2 ratio	Ratio	Raw p-value
H3049D09	Rpo1-1	RNA polymerase 1-1	Unchanged	-0.14617	0.90365	0.09676
H3077D11	Polr2j	Polymerase (RNA) II (DNA directed) polypeptide J	Unchanged	-0.06741	0.95435	0.40850
H3031E04	Rpo1-4	RNA polymerase 1-4	Unchanged	-0.02734	0.98123	0.89045
H3036C03	Polr2a	Polymerase (RNA) II (DNA directed) polypeptide A	Unchanged	0.09411	1.06741	0.77623
H3022B05	Rpo1-3	RNA polymerase 1-3	Unchanged	0.03804	1.02672	0.68630
H3136B02	Polr2g	Polymerase (RNA) II (DNA directed) polypeptide G	Unchanged	0.28678	1.21991	0.42774
H3031C04	Rpo1-3	RNA polymerase 1-3	Unchanged	0.08301	1.05923	0.24530
H3121B06	Rpo1-3	RNA polymerase 1-3	Unchanged	0.06967	1.04947	0.56444

Figure 48 – Genes with in a KEGG Functional Pathway



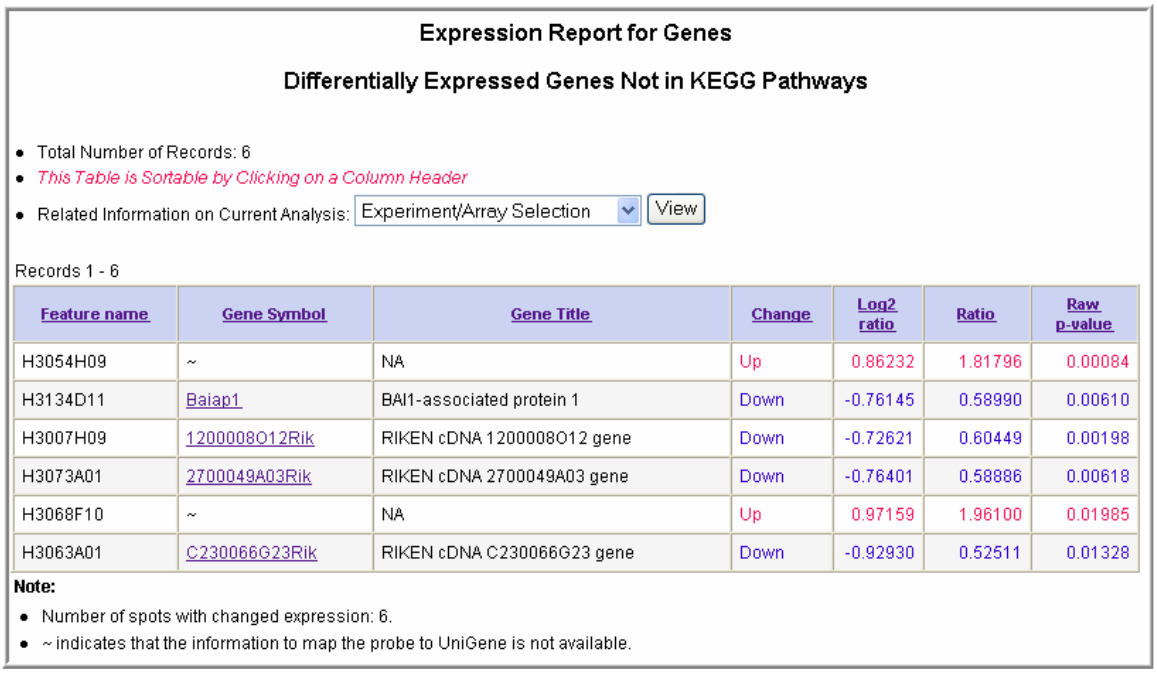
Expression Report for Genes

KEGG Pathway: 1,4-Dichlorobenzene degradation

- Total Number of Records: 7
- *This Table is Sortable by Clicking on a Column Header*
- Related Information on Current Analysis: Experiment/Array Selection View
- Click on the Following Links for Data Export: Export Report Export Sample Data

Feature name	Gene symbol	Gene title	Change	Log2 ratio	Ratio	Raw p-value
H3103E01	Top3a	Topoisomerase (DNA) III alpha	Unchanged	-0.20708	0.86629	0.09018
H3139A05	Top2a	Topoisomerase (DNA) II alpha	Unchanged	0.16958	1.12473	0.81249
H3122G11	Top2b	Topoisomerase (DNA) II beta	Unchanged	0.11681	1.08434	0.66708
H3098D03	Top2b	Topoisomerase (DNA) II beta	Unchanged	-0.06224	0.95777	0.63570
H3023E12	Top3b	Topoisomerase (DNA) III beta	Unchanged	-0.05130	0.96507	0.56209
H3036F09	Top2a	Topoisomerase (DNA) II alpha	Unchanged	0.40787	1.32672	0.65521
H3061C03	Top1	Topoisomerase (DNA) I	Unchanged	0.03482	1.02443	0.48257

Figure 49 - Gene Expression Report for a Selected Pathway



Expression Report for Genes

Differentially Expressed Genes Not in KEGG Pathways

- Total Number of Records: 6
- *This Table is Sortable by Clicking on a Column Header*
- Related Information on Current Analysis: Experiment/Array Selection View

Records 1 - 6

Feature name	Gene Symbol	Gene Title	Change	Log2 ratio	Ratio	Raw p-value
H3054H09	~	NA	Up	0.86232	1.81796	0.00084
H3134D11	Baiap1	BAI1-associated protein 1	Down	-0.76145	0.58990	0.00610
H3007H09	1200008O12Rik	RIKEN cDNA 1200008O12 gene	Down	-0.72621	0.60449	0.00198
H3073A01	2700049A03Rik	RIKEN cDNA 2700049A03 gene	Down	-0.76401	0.58886	0.00618
H3068F10	~	NA	Up	0.97159	1.96100	0.01985
H3063A01	C230066G23Rik	RIKEN cDNA C230066G23 gene	Down	-0.92930	0.52511	0.01328

Note:

- Number of spots with changed expression: 6.
- ~ indicates that the information to map the probe to UniGene is not available.

Figure 50 - Differentially Expressed Genes that are not in any KEGG Pathways

7.9. Perform Gene Category Analysis by Gene Ontology (GO):

When you choose **Perform Gene Category Analysis by Gene Ontology (GO)** (Figure 34), you will enter the menu page for GO analysis (Figure 51).

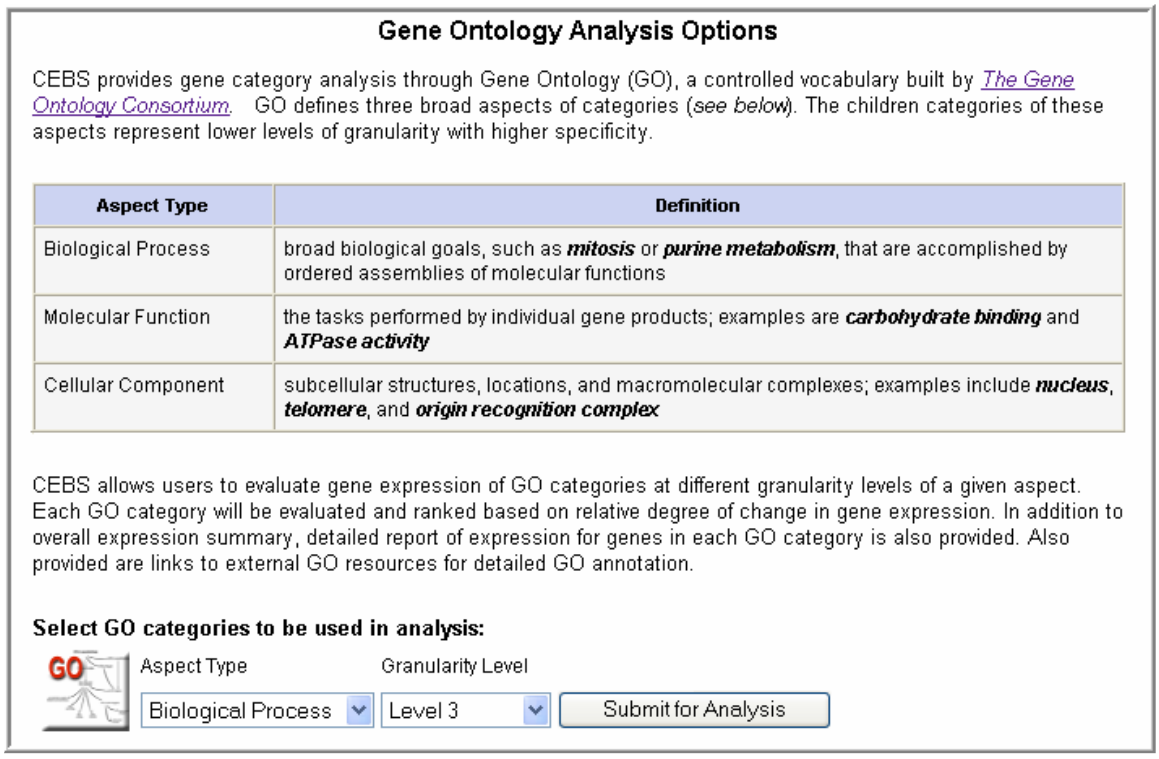


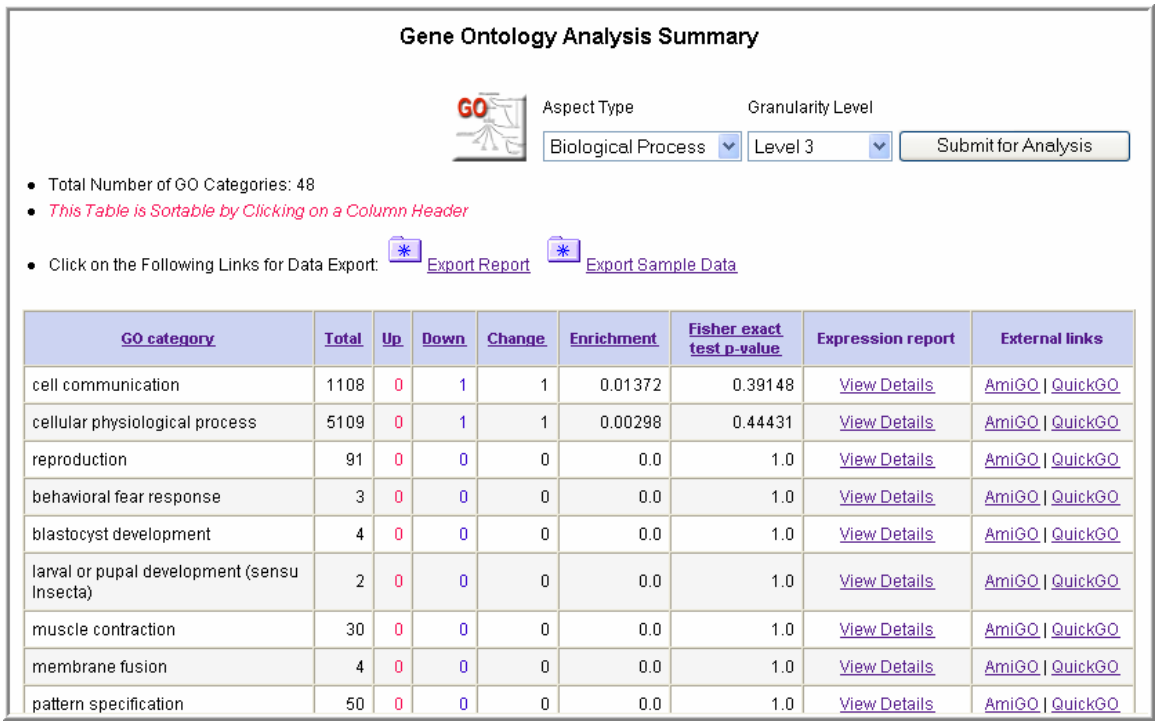
Figure 51 – Gene Ontology Main Page

7.9.1. Perform Expression Analysis Based Upon Gene Ontology

The Gene Ontology Consortium has defined three ways of categorizing genes: by **biological process**, **molecular function** and by **cellular components**. The GO categories have hierarchies based on the classification methods. There are parent and child relationships between GO categories. The parent categories represent more general biological aspects that possess has-a or is-a type of relationship with the children categories, which represent lower level of granularity. CEBS enables users to select GO categories based on the classification methods and level of granularity.

To perform the expression analysis on GO, choose the **Aspect Type** corresponding to one of the classification methods, and **Granularity Level** from the pull-down list on the menu page (Figure 51).

Gene Ontology Analysis Summary: This type of analysis provides the summarized statistical information as well as gene expression data for each of the GO categories that belong to the Aspect Type and Granularity Level specified by the user (Figure 52).



Gene Ontology Analysis Summary

Aspect Type: Biological Process | Granularity Level: Level 3 | Submit for Analysis

- Total Number of GO Categories: 48
- *This Table is Sortable by Clicking on a Column Header*
- Click on the Following Links for Data Export: Export Report Export Sample Data

GO category	Total	Up	Down	Change	Enrichment	Fisher exact test p-value	Expression report	External links
cell communication	1108	0	1	1	0.01372	0.39148	View Details	AmiGO | QuickGO
cellular physiological process	5109	0	1	1	0.00298	0.44431	View Details	AmiGO | QuickGO
reproduction	91	0	0	0	0.0	1.0	View Details	AmiGO | QuickGO
behavioral fear response	3	0	0	0	0.0	1.0	View Details	AmiGO | QuickGO
blastocyst development	4	0	0	0	0.0	1.0	View Details	AmiGO | QuickGO
larval or pupal development (sensu Insecta)	2	0	0	0	0.0	1.0	View Details	AmiGO | QuickGO
muscle contraction	30	0	0	0	0.0	1.0	View Details	AmiGO | QuickGO
membrane fusion	4	0	0	0	0.0	1.0	View Details	AmiGO | QuickGO
pattern specification	50	0	0	0	0.0	1.0	View Details	AmiGO | QuickGO

Figure 52 – Gene Ontology Analysis Summary

GO Category column displays the name of a Gene Ontology term. **Total** column displays the number of all the design elements for which the represented genes are in the specific GO category. **Up**, **Down** and **Change** columns represent the number of up-regulated, down-regulated and changed gene expressions for the comparison criteria you used. The relative proportion of changed genes in a category as compared to overall proportion of changed genes among the total number of genes with available expression value (excluding genes with missing expression data) is presented as an **Enrichment** factor. If the enrichment factor is greater than 1, the gene category has an enrichment of changed compared to all genes, while an enrichment factor less than 1 means the gene category has a depletion in the changed gene. A two-sided **Fisher Exact Test p-value** is also provided as indication of the probability of observing such a number of changed genes in the GO categories by chance.

Each of the GO categories listed is linked to its **Gene Ontology Expression Report** (Figure 53) via the **View Details** hyperlink. In addition, it is also linked to major external GO browsers, **AmiGO** from BDGP and **QuickGO** at EBI, via the GO identifier (GOID) or name. This function allows the users to view definition, relationships between GO categories, as well as detailed description and links to additional information.

Click on **GO Category** column header to sort the GO categories by term name alphabetically. Click on **Total**, **Up**, **Down**, **Change**, **Enrichment** to sort the pathways by the specific column name in descending order. Click on **Fisher Exact Test p-value** to sort the pathways in ascending order.

In addition to these functions, a navigation menu is also provided for the user to view a new set of GO categories by selecting the classification method and level of granularity.

7.9.2. View Expression Report for a Gene Ontology Category

Gene Ontology Expression Report presents the gene-level expression details for the genes on the chip used in the hybridizations of interest that are annotated to a selected GO category (Figure 53). The differentially expressed genes as well as those detected as “unchanged” are presented. CEBS highlights the differentially expressed genes by color-coding the data of up- and down-regulated genes.

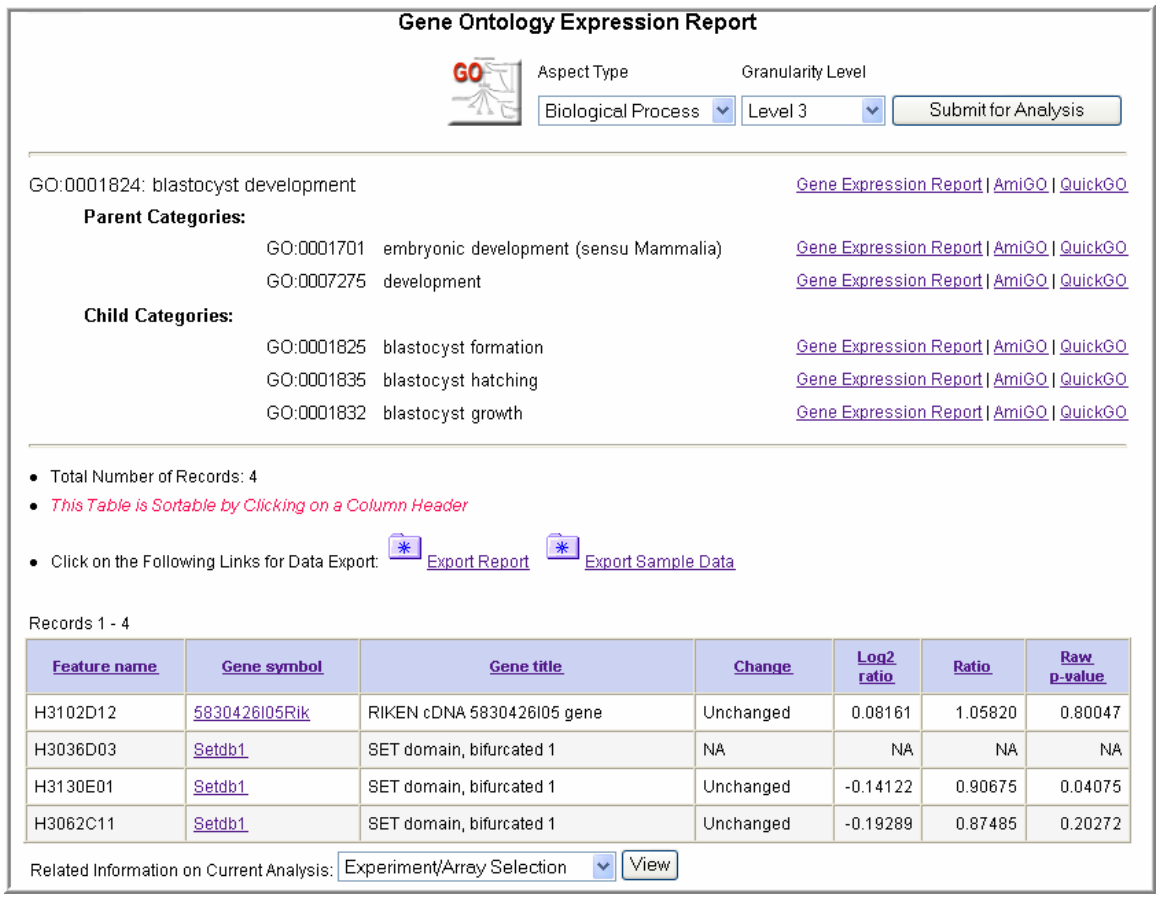


Figure 53 – Gene Ontology Expression Report

For two-channel arrays, the **Feature Name** represents the name of reporter. The expression levels are represented by **log2 ratios** (sample-sample comparisons), or **mean log2 ratios** (group-group comparisons). For Affymetrix arrays, the **Probe Set** represents the name of reporter. **log2 intensities** are used for sample-sample comparisons, while **mean log2 intensities** are used for group-group comparisons. The **Gene Symbol** and **Gene Title** represent the name and title of target gene, respectively. The **Change** column is used to indicate whether the gene expression is up- or down-regulated, by the criteria selected by the user. **log2(A)-log2(B)** represents the difference between (group) A and B, while **Log2 ratio** is used for the difference between two channels. **A/B** represents the fold change between (group) A and B, while **Ratio** is used for the fold change between two channels. A **Raw p-value** is displayed to indicate the confidence level of differential expression between the two groups of arrays, based on the statistical test for the data on the particular gene. An **Adjusted p-value** is derived from the single test p-value, through applying a selected multiple testing procedure to control the false discovery rate or family-wise error rate.

Click on a **Gene Symbol** on the list to view the detailed gene information on **Gene Info Page** (Figure 55).

This report provides links to related GO terms, parent & child terms, to a particular GO category. The users can view the expression of genes in these GO terms via these links to **Gene Ontology Expression Report**. CEBS also links each of these GO terms to major external GO browsers, AmiGO from BDGP and QuickGO at EBI.

A navigation menu is also provided for the users to view a new set of GO categories by selecting the classification method and level of granularity.

Similar to other gene-level expression report, this report table can also be sorted by clicking on the column headers. If the number of genes is large, the report is split into smaller pages with 200 genes per page. Users can navigate through the pages by choosing a page number to go to.

In addition to the gene expression data, the report also provides the related information on current analysis: **Experiment/Array Selection** (Figure 40) and **Options Used for Comparison** (Figure 41). Choose to review the information about the above items by clicking on **View** button.

7.10. View Expression Report for All Differentially Expressed Genes:

Select **View Expression Report for All Differentially Expressed Genes** option (Figure 34) to view **All Differentially Expressed Genes** report (Figure 54), which displays all differentially expressed genes represented in the chip based on the criteria you specified in the comparison analysis.

All Differentially Expressed Genes

- Total Number of Differentially Expressed Genes: 396
- *This Table is Sortable by Clicking on a Column Header*
- Related Information on Current Analysis: Experiment/Array Selection View
- Click on the Following Links for Data Export: Export Report Export Sample Data

Records 1 - 200 Page 1 of 2 Go to Page 1

Feature name	Gene symbol	Gene title	Change	Log2 ratio	Ratio	Raw p-value
H3083A01	Ppp1cb	Protein phosphatase 1, catalytic subunit, beta isoform	Down	-0.39942	0.75816	0.02962
H3083A09	~	NA	Down	-0.34531	0.78714	0.03358
H3099A03	Eif4enif1	Eukaryotic translation initiation factor 4E nuclear import factor 1	Down	-0.30973	0.80679	0.00869
H3103A07	Gpr34	G protein-coupled receptor 34	Down	-0.27740	0.82507	0.01083
H3103E09	Slc37a3	Solute carrier family 37 (glycerol-3-phosphate transporter), member 3	Down	-0.26716	0.83095	0.00063
H3103E11	Unnamed Gene	Transcribed locus	Down	-0.33876	0.79072	0.04295
H3155E07	Dock4	Dedicator of cytokinesis 4	Down	-0.31432	0.80423	0.04496
H3012E09	~	NA	Down	-0.34058	0.78973	0.01976
H3020E11	Ccnf	Cyclin F	Down	-0.45134	0.73136	0.03778
H3006B08	D14Abb1e	DNA segment, Chr 14, Abbott 1 expressed	Down	-0.30935	0.80700	0.03834
H3010F02	Lmln	Leishmanolysin-like (metallopeptidase M8 family)	Down	-0.27998	0.82360	0.01144
H3014B10	Ncl	Nucleolin	Up	0.29443	1.22640	0.02434
H3014F12	2310061A22Rik	RIKEN cDNA 2310061A22 gene	Up	0.28337	1.21703	0.02937

Figure 54 – All Differentially Expressed Genes

For two-channel arrays, the **Feature Name** represents the name of reporter. The expression levels are represented by **log2 ratios** (sample-sample comparisons), or **mean log2 ratios** (group-group comparisons). For Affymetrix arrays, the **Probe Set** represents the name of reporter. **log2 intensities** are used for sample-sample comparisons, while **mean log2 intensities** are used for group-group comparisons. The **Gene Symbol** and **Gene Title** represent the name and

title of target gene, respectively. The **Change** column is used to indicate whether the gene expression is up- or down-regulated, by the criteria selected by the user. **log2(A)-log2(B)** represents the difference between (group) A and B, while **Log2 ratio** is used for the difference between two channels. **A**/**B** represents the fold change between (group) A and B, while **Ratio** is used for the fold change between two channels. A **Raw p-value** is displayed to indicate the confidence level of differential expression between the two groups of arrays, based on the statistical test for the data on the particular gene. An **Adjusted p-value** is derived from the single test p-value, through applying a selected multiple testing procedure to control the false discovery rate or family-wise error rate.

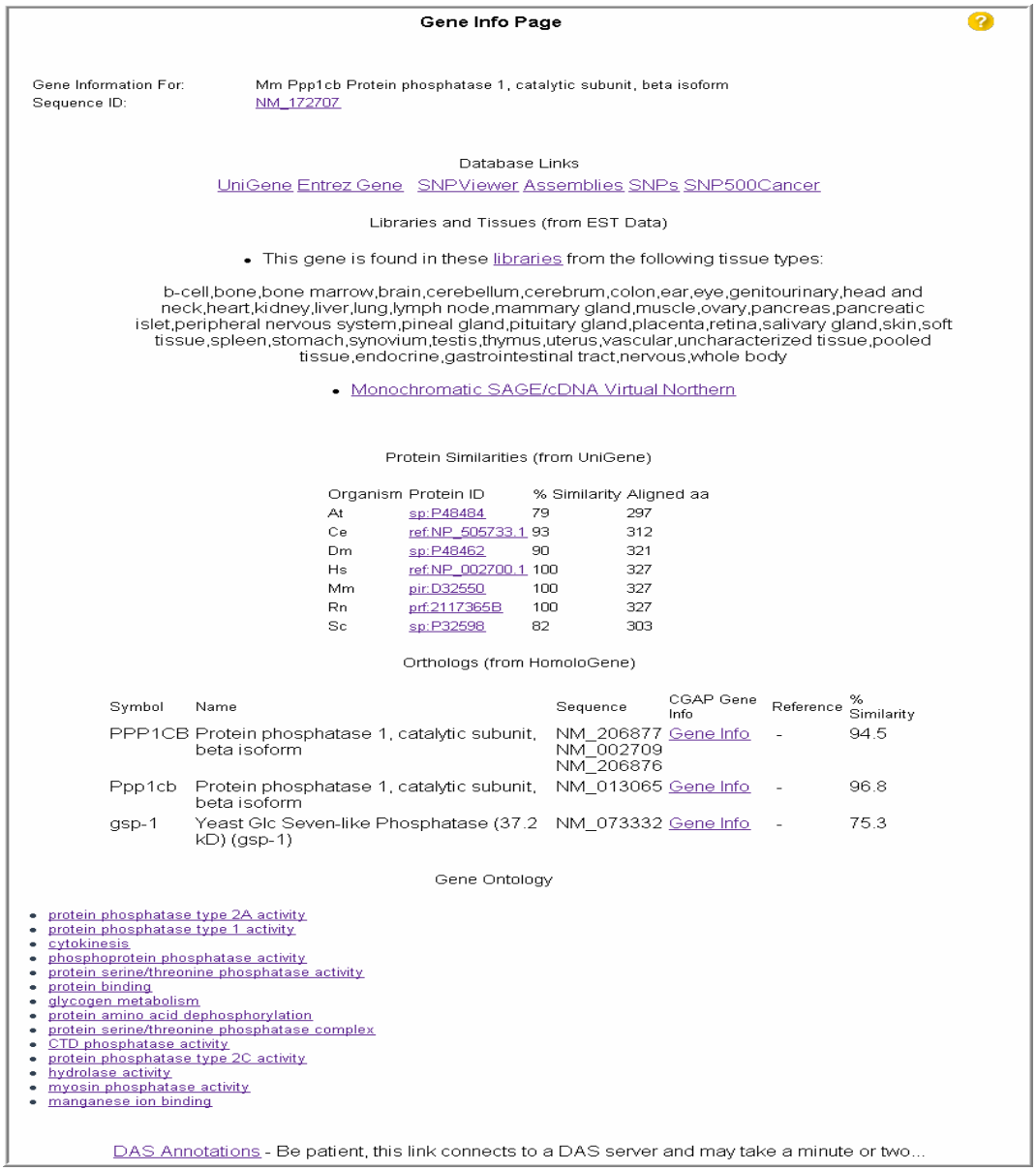
Click on the name of a **Gene Symbol** (Figure 54) to view the detailed gene information on **Gene Info Page** (Figure 55).

Similar to other gene level expression report, this report table can also be sorted by clicking on the column headers. If the number of genes is large, the report is split into smaller pages with 200 genes per page. Users can navigate through the pages by choosing a page number to go to.

In addition to the gene expression data, the report also provides the related information on current analysis: **Experiment**/**Array Selection** (Figure 40) and **Options Used for Comparison** (Figure 41). Choose to review the information about these items by clicking on **View** button.

Gene Info Page

Gene Info Page presents basic information about the gene and links to biological and functional information on the gene you selected (Figure 55), including gene annotation from UniGene and LocusLink, mRNA and EST sequences information, protein sequence similarity information from UniGene, library data (SAGE, EST) from CGAP, SNP information from CGAP, and Gene Ontology information from *The Gene Ontology Consortium*. A link to DAS Annotations is also displayed to connect to DAS server. Click on a link to view the corresponding detailed information.



Gene Info Page

Gene Information For: Mm Ppp1cb Protein phosphatase 1, catalytic subunit, beta isoform
Sequence ID: NM_172707

Database Links
UniGene Entrez Gene SNPViewer Assemblies SNPs SNP500Cancer

Libraries and Tissues (from EST Data)

- This gene is found in these libraries from the following tissue types:

b-cell,bone,bone marrow,brain,cerebellum,cerebrum,colon,ear,eye,genitourinary,head and neck,heart,kidney,liver,lung,lymph node,mammary gland,muscle,ovary,pancreas,pancreatic islet,peripheral nervous system,pineal gland,pituitary gland,placenta,retina,salivary gland,skin,soft tissue,spleen,stomach,synovium,testis,thymus,uterus,vascular,uncharacterized tissue,pooled tissue,endocrine,gastrointestinal tract,nervous,whole body

- Monochromatic SAGE/cDNA Virtual Northern

Protein Similarities (from UniGene)

Organism	Protein ID	% Similarity	Aligned aa
At	sp:P48484	79	297
Ce	ref:NP_505733.1	93	312
Dm	sp:P48462	90	321
Hs	ref:NP_002700.1	100	327
Mm	pir:D32550	100	327
Rn	prf:2117365B	100	327
Sc	sp:P32598	82	303

Orthologs (from HomoloGene)

Symbol	Name	Sequence	CGAP Gene Info	Reference	% Similarity
PPP1CB	Protein phosphatase 1, catalytic subunit, beta isoform	NM_206877 NM_002709 NM_206876	Gene Info	-	94.5
Ppp1cb	Protein phosphatase 1, catalytic subunit, beta isoform	NM_013065	Gene Info	-	96.8
gsp-1	Yeast Glc Seven-like Phosphatase (37.2 kD) (gsp-1)	NM_073332	Gene Info	-	75.3

Gene Ontology

- protein phosphatase type 2A activity
- protein phosphatase type 1 activity
- cytokinesis
- phosphoprotein phosphatase activity
- protein serine/threonine phosphatase activity
- protein binding
- glycogen metabolism
- protein amino acid dephosphorylation
- protein serine/threonine phosphatase complex
- CTD phosphatase activity
- protein phosphatase type 2C activity
- hydrolase activity
- myosin phosphatase activity
- manganese ion binding

DAS Annotations - Be patient, this link connects to a DAS server and may take a minute or two...

Figure 55 – Gene Info Page

8. Quality Control Information

CEBS provides quality control and statistical information for each array in the experiment. To view the quality control and statistical information, search experiment(s), and use **Click to View** for **Data Quality Information** in the **Detailed View** of **Experiment Information Report** page (Figure 56).

Experiment Information Report

Brief View

- Click on "Experiment ID" to view experiment information report in the Detailed View.
- For analysis, use the check box below to select experiment(s), then click on "Analyze Selected Experiment(s)".
- Multiple experiments can be analyzed together, if they have same platform & matched array design (ID).

Analyze	Experiment ID	Investigator	Experiment Title	Image Processing Software	Visibility	Array Design (ID)
☐	522398544	Alexandra Heinloth	Gene Expression Profiling of F344/N Rat Livers After Acute Acetaminophen Exposure	Affymetrix	Public	Rat230_2 (346482001)

Analyze Selected Experiment(s)

Detailed View

Gene Expression Profiling of F344/N Rat Livers After Acute Acetaminophen Exposure

Experiment ID	522398544
Investigator Name	Alexandra Heinloth
Organization	NCT
Experiment Type	Treatment vs. Untreated Comparison
Species	[Rat]
Tissue(s)	[Liver Left Lateral Lobe]
Image Processing Software	Affymetrix
Array Design Name	Rat230_2
Array Design ID	346482001
Stressor Name(s)	Acetaminophen
Experimental Variable(s)	Acetaminophen (Dose Level,Time)
Characteristics That Vary Between Samples	None
Publication	
Submission Date	2004-12-15
Experiment Description	Microarray analysis of liver RNA from F344/N rats after acute exposure to Acetaminophen (50mg/kg, 150mg/kg, 1500mg/kg) on RAT230_2 Affymetrix arrays
Data Files	Click to Download
Data Quality Information	Click to View

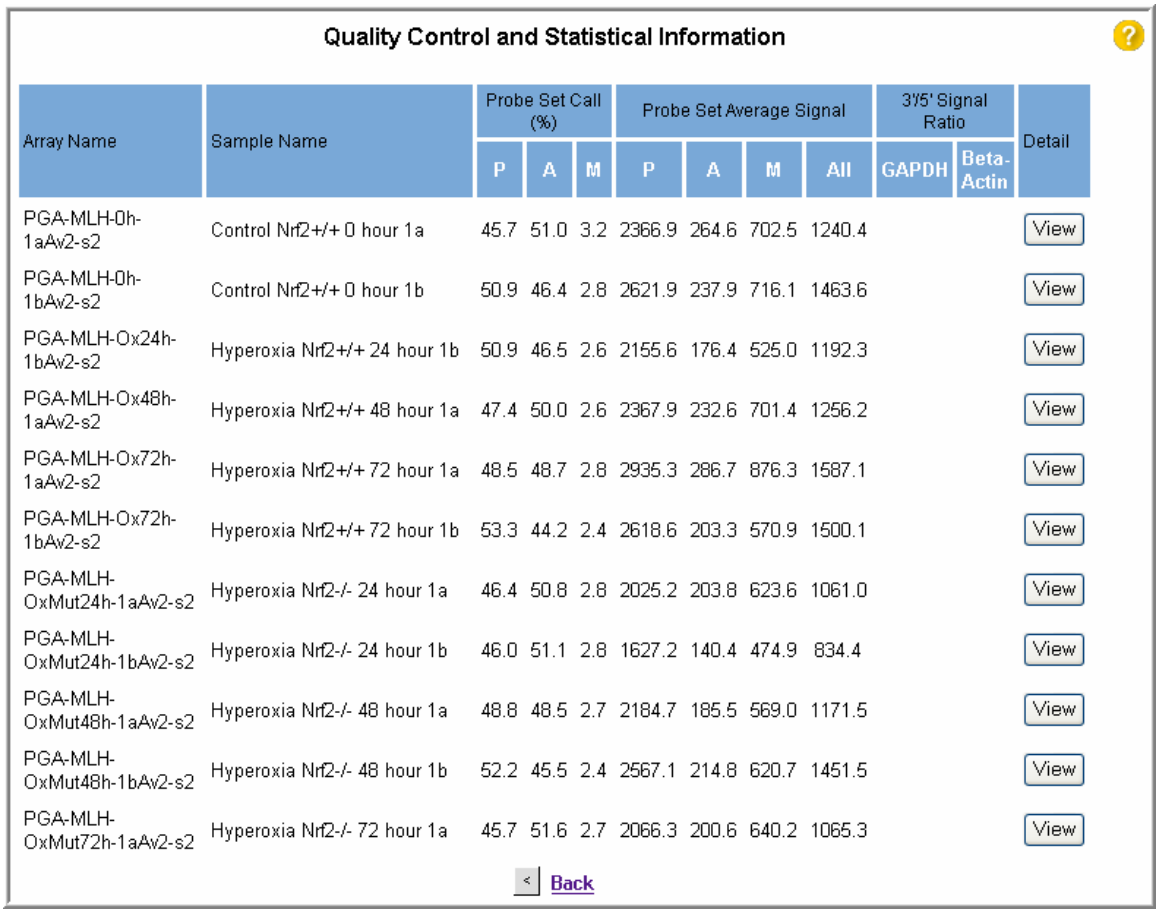
GO TOP

Figure 56 – Experiment Information Report

The displayed information varies among different image processing software. Currently the quality control information can be displayed for the experiment data from Affymetrix, Agilent, GenePix and ArraySuite image processing software.

8.1. Affymetrix

The quality control information is retrieved from the .rpt files from Affymetrix MAS 5 software. The information is displayed for each array in the experiment. The quality control and statistical information includes percentage of probe sets that are called as Present (P), Absent (A) and Marginal (M); the average signal for Present (P), Absent (A), Marginal (M) and all the probe set; the 3'/5' signal ratios for housekeeping genes i.e. GAPDH and Beta-actin if they are available in the chip. Place the computer mouse over the column name to view the description (Figure 57). Click **View** button to view detailed quality control information (Figure 58). The information in the detail page includes software parameter setting, background and noise, number of probe sets for each type of calls and signal intensity, housekeeping and spike control genes information. For more information, please review **Affymetrix's GeneChip Expression Analysis Technical Manual**.



Quality Control and Statistical Information

Array Name	Sample Name	Probe Set Call (%)			Probe Set Average Signal				3'/5' Signal Ratio		Detail
		P	A	M	P	A	M	All	GAPDH	Beta-Actin	
PGA-MLH-0h-1aAv2-s2	Control Nrf2+/+ 0 hour 1a	45.7	51.0	3.2	2366.9	264.6	702.5	1240.4			View
PGA-MLH-0h-1bAv2-s2	Control Nrf2+/+ 0 hour 1b	50.9	46.4	2.8	2621.9	237.9	716.1	1463.6			View
PGA-MLH-Ox24h-1bAv2-s2	Hyperoxia Nrf2+/+ 24 hour 1b	50.9	46.5	2.6	2155.6	176.4	525.0	1192.3			View
PGA-MLH-Ox48h-1aAv2-s2	Hyperoxia Nrf2+/+ 48 hour 1a	47.4	50.0	2.6	2367.9	232.6	701.4	1256.2			View
PGA-MLH-Ox72h-1aAv2-s2	Hyperoxia Nrf2+/+ 72 hour 1a	48.5	48.7	2.8	2935.3	286.7	876.3	1587.1			View
PGA-MLH-Ox72h-1bAv2-s2	Hyperoxia Nrf2+/+ 72 hour 1b	53.3	44.2	2.4	2618.6	203.3	570.9	1500.1			View
PGA-MLH-OxMut24h-1aAv2-s2	Hyperoxia Nrf2-/- 24 hour 1a	46.4	50.8	2.8	2025.2	203.8	623.6	1061.0			View
PGA-MLH-OxMut24h-1bAv2-s2	Hyperoxia Nrf2-/- 24 hour 1b	46.0	51.1	2.8	1627.2	140.4	474.9	834.4			View
PGA-MLH-OxMut48h-1aAv2-s2	Hyperoxia Nrf2-/- 48 hour 1a	48.8	48.5	2.7	2184.7	185.5	569.0	1171.5			View
PGA-MLH-OxMut48h-1bAv2-s2	Hyperoxia Nrf2-/- 48 hour 1b	52.2	45.5	2.4	2567.1	214.8	620.7	1451.5			View
PGA-MLH-OxMut72h-1aAv2-s2	Hyperoxia Nrf2-/- 72 hour 1a	45.7	51.6	2.7	2066.3	200.6	640.2	1065.3			View

< Back

Figure 57 – Quality Control & Statistical Information for Affymetrix Arrays

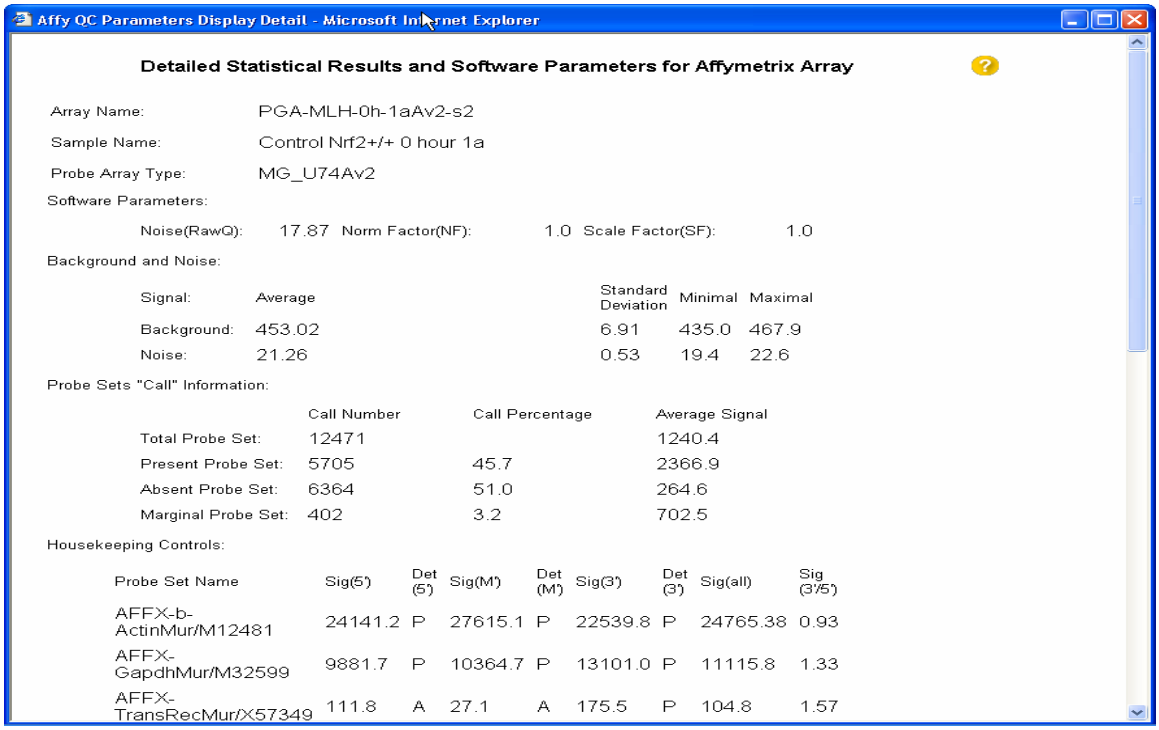
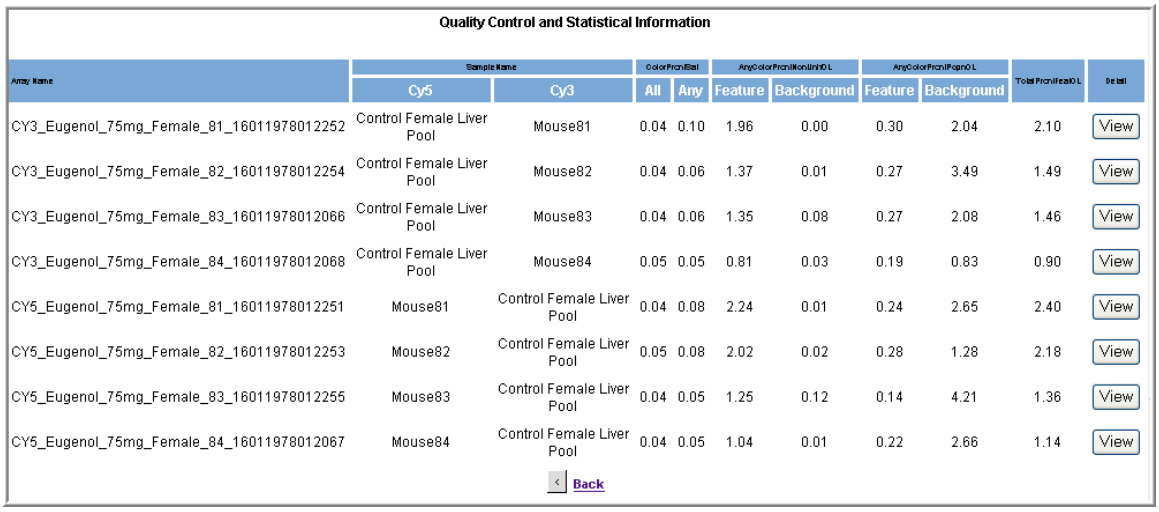


Figure 58 – Detailed Statistical Results & Software Parameters for Affymetrix Arrays

8.2. Agilent

The quality control information is retrieved from the uploaded Agilent data file. In the **Quality Control and Statistical Information** page, the several most important statistical results are displayed. Please put computer mouse over the column name to view the description (Figure 59). Click **View** button to view more quality control information (Figure 60). In the **Detailed Statistical Results for Agilent Array** page, detailed quality control information retrieved from the statistical result section of the data file for each array is displayed. For more information, please review user manual from Agilent's microarray feature extraction software.

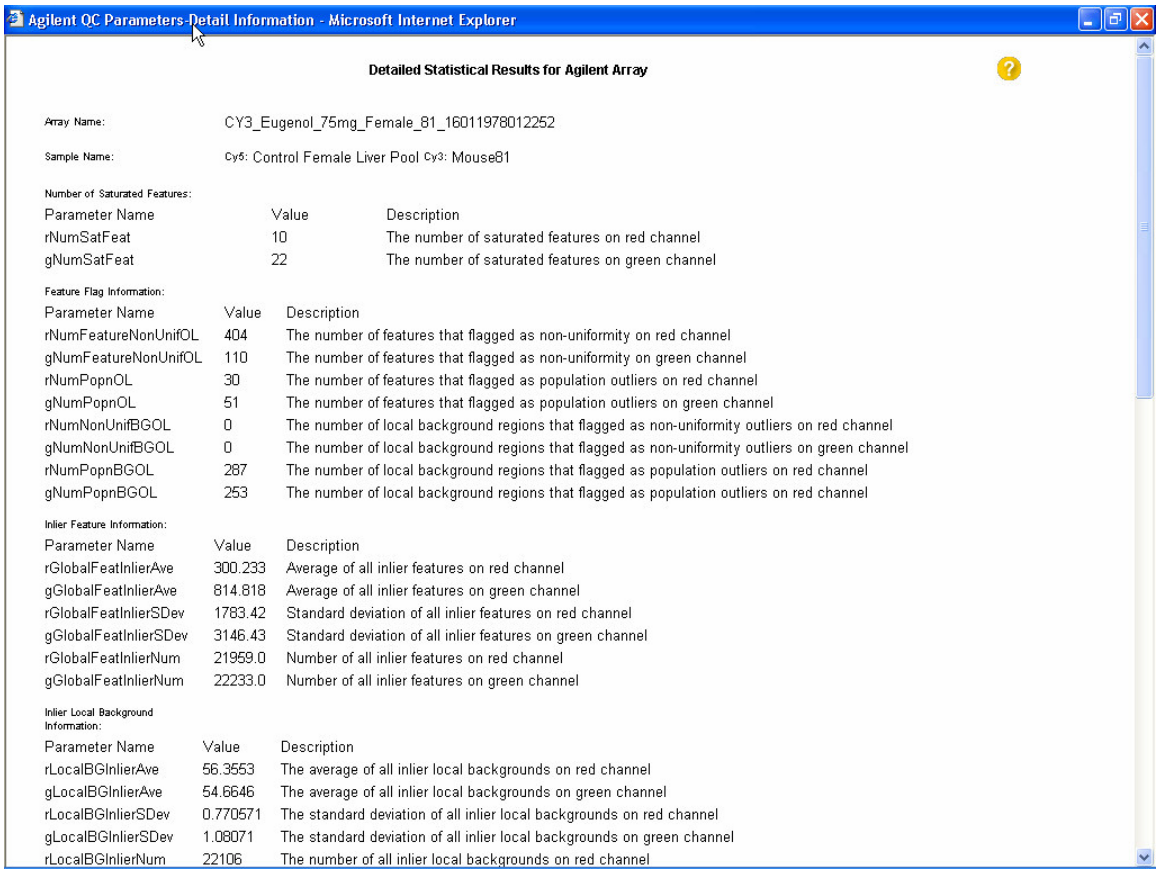


Quality Control and Statistical Information

Array Name	Sample Name		ColorPrcntBad		AnyColorPrcntNonUnifOL		AnyColorPrcntPopnOL		Total PrcntFeatOL	Detail
	Cy5	Cy3	All	Any	Feature	Background	Feature	Background		
CY3_Eugenol_75mg_Female_81_16011978012252	Control Female Liver Pool	Mouse81	0.04	0.10	1.96	0.00	0.30	2.04	2.10	View
CY3_Eugenol_75mg_Female_82_16011978012254	Control Female Liver Pool	Mouse82	0.04	0.06	1.37	0.01	0.27	3.49	1.49	View
CY3_Eugenol_75mg_Female_83_16011978012066	Control Female Liver Pool	Mouse83	0.04	0.06	1.35	0.08	0.27	2.08	1.46	View
CY3_Eugenol_75mg_Female_84_16011978012068	Control Female Liver Pool	Mouse84	0.05	0.05	0.81	0.03	0.19	0.83	0.90	View
CY5_Eugenol_75mg_Female_81_16011978012251	Mouse81	Control Female Liver Pool	0.04	0.08	2.24	0.01	0.24	2.65	2.40	View
CY5_Eugenol_75mg_Female_82_16011978012253	Mouse82	Control Female Liver Pool	0.05	0.08	2.02	0.02	0.28	1.28	2.18	View
CY5_Eugenol_75mg_Female_83_16011978012255	Mouse83	Control Female Liver Pool	0.04	0.05	1.25	0.12	0.14	4.21	1.36	View
CY5_Eugenol_75mg_Female_84_16011978012067	Mouse84	Control Female Liver Pool	0.04	0.05	1.04	0.01	0.22	2.66	1.14	View

Back

Figure 59 – Quality Control & Statistical Information for Agilent Arrays



Agilent QC Parameters-Detail Information - Microsoft Internet Explorer

Detailed Statistical Results for Agilent Array

Array Name: CY3_Eugenol_75mg_Female_81_16011978012252

Sample Name: Cy5: Control Female Liver Pool Cy3: Mouse81

Number of Saturated Features:

Parameter Name	Value	Description
rNumSatFeat	10	The number of saturated features on red channel
gNumSatFeat	22	The number of saturated features on green channel

Feature Flag Information:

Parameter Name	Value	Description
rNumFeatureNonUnifOL	404	The number of features that flagged as non-uniformity on red channel
gNumFeatureNonUnifOL	110	The number of features that flagged as non-uniformity on green channel
rNumPopnOL	30	The number of features that flagged as population outliers on red channel
gNumPopnOL	51	The number of features that flagged as population outliers on green channel
rNumNonUnifBGOL	0	The number of local background regions that flagged as non-uniformity outliers on red channel
gNumNonUnifBGOL	0	The number of local background regions that flagged as non-uniformity outliers on green channel
rNumPopnBGOL	287	The number of local background regions that flagged as population outliers on red channel
gNumPopnBGOL	253	The number of local background regions that flagged as population outliers on red channel

Inlier Feature Information:

Parameter Name	Value	Description
rGlobalFeatInlierAve	300.233	Average of all inlier features on red channel
gGlobalFeatInlierAve	814.818	Average of all inlier features on green channel
rGlobalFeatInlierSDev	1783.42	Standard deviation of all inlier features on red channel
gGlobalFeatInlierSDev	3146.43	Standard deviation of all inlier features on green channel
rGlobalFeatInlierNum	21959.0	Number of all inlier features on red channel
gGlobalFeatInlierNum	22233.0	Number of all inlier features on green channel

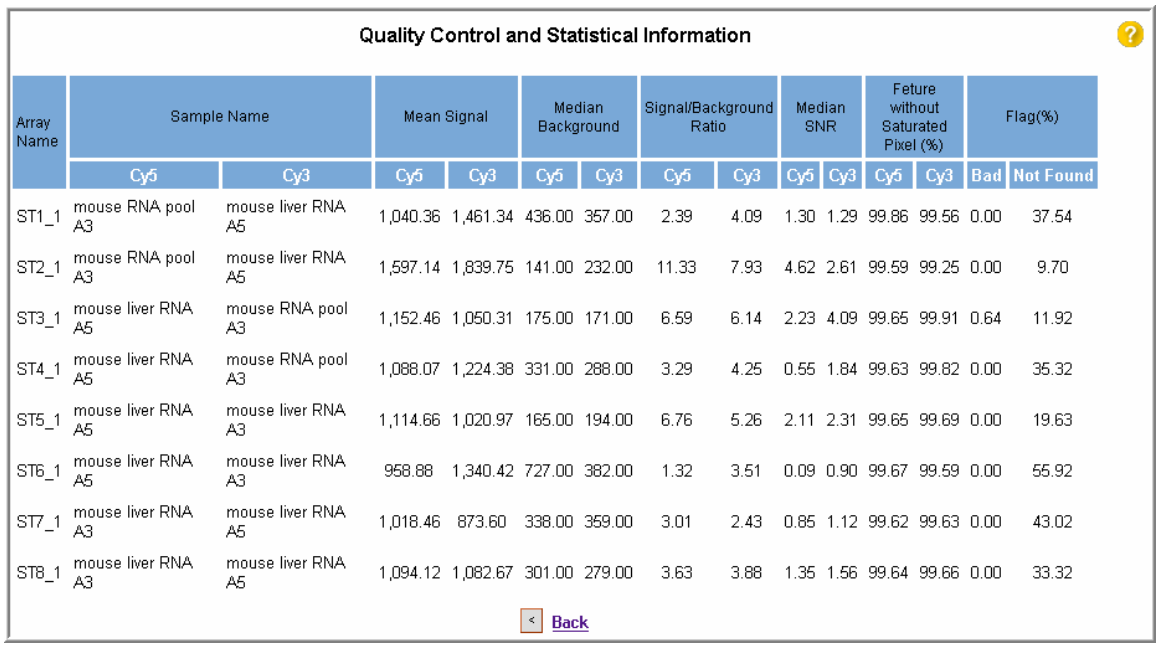
Inlier Local Background Information:

Parameter Name	Value	Description
rLocalBGInlierAve	56.3553	The average of all inlier local backgrounds on red channel
gLocalBGInlierAve	54.6646	The average of all inlier local backgrounds on green channel
rLocalBGInlierSDev	0.770571	The standard deviation of all inlier local backgrounds on red channel
gLocalBGInlierSDev	1.08071	The standard deviation of all inlier local backgrounds on green channel
rLocalBGInlierNum	22106	The number of all inlier local backgrounds on red channel

Figure 60 - Detailed Statistical Results & Software Parameters for Agilent Arrays

8.3. GenePix

a. The quality control information is computed from the output of GenePix image processing software. It provides summarized array level quality control information (Figure 61).
b. “Mean Signal” is the mean signal intensity value for all the features in the array. For Cy5 channel, signal intensity of each feature is computed from “F635 Mean” minus “B635 Mean” of the output file, then “Mean Signal” for Cy5 is computed from signal intensity values of all the features in the array. For Cy3 channel, signal intensity of each feature is computed from “F532 Mean” minus “B532 Mean” of the data file, then “Mean Signal” for Cy3 is computed from signal intensity values of all the features in the array.
c. “Median Background” is the median background value for all features in the array. For Cy5 channel, the “Median Background” is computed from "B635 Mean" values of all the features. For Cy3 channel, the “Median Background” is computed from "B532 Mean" values of all the features.
d. “Signal/Background” is the ratio of Mean Signal / Mean Background. The ratio of Cy5 channel is computed from the values of “Mean Signal” and “Median Background” from the cy5 channel. The ratio of Cy3 channel is computed from the values of “Mean Signal” and “Median Background” ” from the cy3 channel.
e. “Median SNR” is the median of the SNR (Signal-to-Noise) values of all the features in the array. For Cy5 channel, the SNR for each feature is computed from the formula of (F635 Median - B635 Median) / B635 SD. The median value of Cy5 channel is then computed from SNR values of all the features in the array. For Cy3 channel, the SNR for each feature is computed from the formula of (F532 Median - B532 Median) / B532 SD. The median value of Cy3 channel is then computed from SNR values of all the features in the array.
f. “Feature without Saturated Pixels (%)”: The "F635 % Sat" and "F532% sat" in the data file indicates the percentage of the feature pixels that are saturated in Cy5 channel and Cy3 channel respectively. “Feature without Saturated Pixels (%)” is the percentage of the features with values of zero for “F635 % Sat” (Cy5) or "F532% sat" (Cy3) in the array.
g. “Flag (%)” are the percentages of the features that are indicated as “Bad” (-100) or “Not Found” (-50) in the “Flag” column in the .gpr file for an array.



Quality Control and Statistical Information

Array Name	Sample Name		Mean Signal		Median Background		Signal/Background Ratio		Median SNR		Feture without Saturated Pixel (%)		Flag(%)	
	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5	Cy3	Bad	Not Found
ST1_1	mouse RNA pool A3	mouse liver RNA A5	1,040.36	1,461.34	436.00	357.00	2.39	4.09	1.30	1.29	99.86	99.56	0.00	37.54
ST2_1	mouse RNA pool A3	mouse liver RNA A5	1,597.14	1,839.75	141.00	232.00	11.33	7.93	4.62	2.61	99.59	99.25	0.00	9.70
ST3_1	mouse liver RNA A5	mouse RNA pool A3	1,152.46	1,050.31	175.00	171.00	6.59	6.14	2.23	4.09	99.65	99.91	0.64	11.92
ST4_1	mouse liver RNA A5	mouse RNA pool A3	1,088.07	1,224.38	331.00	288.00	3.29	4.25	0.55	1.84	99.63	99.82	0.00	35.32
ST5_1	mouse liver RNA A5	mouse liver RNA A3	1,114.66	1,020.97	165.00	194.00	6.76	5.26	2.11	2.31	99.65	99.69	0.00	19.63
ST6_1	mouse liver RNA A5	mouse liver RNA A3	958.88	1,340.42	727.00	382.00	1.32	3.51	0.09	0.90	99.67	99.59	0.00	55.92
ST7_1	mouse liver RNA A3	mouse liver RNA A5	1,018.46	873.60	338.00	359.00	3.01	2.43	0.85	1.12	99.62	99.63	0.00	43.02
ST8_1	mouse liver RNA A3	mouse liver RNA A5	1,094.12	1,082.67	301.00	279.00	3.63	3.88	1.35	1.56	99.64	99.66	0.00	33.32

Back

Figure 61 - Quality Control & Statistical Information for GenePix Arrays

8.4. ArraySuite

The quality control information is computed from the output of Array Suite image software. This is summarized array level quality control information (Figure 62).

a. “Mean Signal” is the mean value that is computed from all features’ “SR_Mean” or “SG_Mean” values of the output file. For the sample channel, the “Mean Signal” is computed from feature’s “SR_Mean” in the array; for the reference channel, the “Mean Signal” is computed from feature’s “SG_Mean” in the array.

b. “Median Background” is the median value that is computed from all feature’s mean background intensity values. For the sample channel, the “Median Background” is computed from feature’s "SR_bkMean" in the array; for reference channel, “Median Background” is computed from feature’s "SG_bkMean" in the array.

c. “Signal/Background” is a ratio of “Mean Signal” / “Mean Background”. The calculations for “Mean Signal” and “Median Background” values are described above. The “Signal/Background” ratio for the sample channel is computed from the “Mean Signal” and “Median Background” sample values. The “Signal/Background” ratio for the reference channel is computed from the “Mean Signal” and “Median Background” reference values.

d. “Median SNR” is the median value of all the features’ Signal-to-Noise values for either sample or reference channel in the array. The “Median SNR” for sample channel is computed from feature’s

"SR_SNR" values in the array; the "Median SNR" for reference channel is computed from feature's "SG_SNR" values in the array.

e. "Feature with Good Quality (%)": In the ArraySuite output file, "SR_iQuality" and "SG_iQuality are used to indicate whether a specific feature has good image qualities for sample and reference channels. The "Feature with Good Quality (%)" is the percentage of the features with that SR_iQuality or SG_iQuality value of 1 for each channel, respectively.

Quality Control and Statistical Information

Array Name	Sample Name		Mean Signal		Median Background		Signal/Background		Median SNR		Feature with Good Quality (%)	
	Sample	Reference	Sample	Reference	Sample	Reference	Sample	Reference	Sample	Reference	Sample	Reference
p90s35_si_fci	mouse liver RNA A3 (SR)-Cy3	mouse liver RNA A5 (SG)-Cy5	1,479.43	2,215.17	382.80	363.70	3.86	6.09	165.50	294.45	97.77	99.05
p90s36_si_fci	mouse liver RNA A3 (SR)-Cy3	mouse liver RNA A5 (SG)-Cy5	1,390.64	2,048.87	379.90	361.80	3.66	5.66	151.35	256.75	97.82	99.01
p90s37_si_fci	mouse liver RNA A3 (SR)-Cy5	mouse liver RNA A5 (SG)-Cy3	1,798.33	1,143.57	361.80	380.60	4.97	3.00	242.25	125.90	98.30	99.10
p90s38_si_fci	mouse liver RNA A3 (SR)-Cy5	mouse liver RNA A5 (SG)-Cy3	1,641.75	1,544.64	350.60	371.60	4.68	4.16	363.35	229.50	98.51	99.00
p90s39_si_fci	mouse RNA pool A3 (SR)-Cy3	mouse liver RNA A5 (SG)-Cy5	1,341.04	1,912.79	379.10	360.00	3.54	5.31	165.35	280.95	97.31	99.09
p90s41_si_fci	mouse RNA pool A3 (SR)-Cy5	mouse liver RNA A5 (SG)-Cy3	1,845.51	1,642.32	350.60	372.00	5.26	4.41	408.70	244.15	98.48	99.07
p90s42_si_fci	mouse RNA pool A3 (SR)-Cy5	mouse liver RNA A5 (SG)-Cy3	1,903.20	2,048.56	350.60	371.10	5.43	5.52	421.15	286.70	98.18	98.61
p90s43_si_fci	mouse RNA pool A3 (SR)-Cy3	mouse liver RNA A5 (SG)-Cy5	2,527.29	1,479.45	359.00	343.60	7.04	4.31	442.60	345.45	97.88	98.95

Back

Figure 62 - Quality Control & Statistical Information for ArraySuite Arrays